University of Alberta

Statistical Issues with Repeated Measurement Designs

By

Rong Huang ©

A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in

Statistics

Department of Mathematical and Statistical Sciences

Edmonton, Alberta

Fall 2001

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-68948-4

Canada

University of Alberta

Library Release Form

Name of Author: Rong Huang

Title of Thesis: Statistical Issues with Repeated Measurement Designs

Degree: Doctor of Philosophy

Year this Degree Granted: 2001

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Rong Huang

5119-106A st.
Edmonton, AB, T6H 2W6

Date: Sep. 21, 2001

To my parents, Yiquan Huang and Yunlan Ji

UNIVERSITY OF ALBERTA

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Statistical Issues with Repeated Measurement Designs** submitted by **Rong Huang** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in Statistics.

Peter Hooper

Dr. Peter Hooper (Chair)

K. C. Carriere

Dr. K. C. Carriere (Supervisor)

Dr. Rohana Karunamuni

Stephen Newman

Dr. Stephen Newman (Public Health Sciences)

Dr. Richard Cook (University of Waterloo)

Date: Sep 7, 2001

ABSTRACT

This thesis is a study of design and analysis issues arising from repeated measurement designs. We have constructed repeated measurement designs under much more general models than those constructed in previous studies, and have suggested new alternative approaches to the analysis of repeated measures data with missing values.

We approached the construction of repeated measurement designs in two different ways: (1) constructing optimal designs with the assumption that the model includes random subject effects and that the serial correlation structure for measurement error is autoregressive; (2) constructing adaptive designs that allow the model to have an unknown error structure with heterogeneity. In the first approach, we constructed optimal designs by maximizing the corresponding information on treatment effects, noting some similarities and differences as compared to the optimal designs constructed under less general models. In the second approach, we constructed adaptive designs using updated information from the available data in sequence. The allocation rules for assigning subjects to sequences were derived in such a way as to maximize the increment of the information on treatment effects obtained from new subjects. These allocation rules were adapted to account for any loss of information caused by incorrectly specified error structures. We used a simulation study to compare the efficiency of the adaptive designs with the designs constructed without accommodating

the subject heterogeneity.

In the analysis of the repeated measures data with missing values, we also considered two different approaches to make up "complete" data sets: (1) using proxy information for missing data, and (2) using a multiple imputation strategy to fill in for missing values. We developed small-sample testing procedures for both approaches, under assumptions that the repeated measurements are from a multivariate normal distribution and the missing data occur at random. Simulation studies indicated that our proposed procedures performed well for small-sample repeated measures data. We concluded that the use of proxy information is more powerful than the other existing incomplete data analyses methods. Lastly, we found that, in general, the multiple imputation method is not a worthwhile strategy, especially in situations where proxy data are obtainable or available.

ACKNOWLEDGEMENT

I am extremely grateful to Professor K.C. Carriere for her supervision, first-year summer support and assistance with my successful applications for an Alberta Heritage Foundation for Medical Research studentship and the Izaack Walton Killam Memorial Scholarship. Professor Carriere also supported my participation in the International Conference on Health Policy Research in Washington, D.C., the International Biometrics Society meetings in San Francisco, and the Joint Meeting of Statistics Society of Canada in Burnaby, B.C.

I thank Dr. Peter Hooper, Dr. Rohan Karunamuni, Dr. Richard Cook, and Dr. Stephen Newman, the members of my supervisory and examining committee for reading my thesis carefully with many valuable comments. I also thank the Department of Mathematical and Statistical Sciences, University of Alberta, for having granted me a teaching assistantship and sessional lectureship— both precious opportunities to improve my teaching skills. Thanks also go to the Trustee of the estate of Killam and Alberta Heritage Foundation for Medical Research for awarding me prestigious scholarships, thereby giving me time to focus on writing my thesis.

Finally, I wish to thank my family. Their unfailing support and encouragement, and their love, have made me strong enough and courageous enough to pursue my goal to completion. Any honor or glory that I have achieved, or may achieve, belongs to them.

Contents

1 Introduction 1

1.1 Background of the thesis 1

1.1.1 Design issues 2

1.1.2 Analysis issues 6

1.2 Thesis overview 9

2 Optimal Two-Treatment Repeated Measurement Designs 11

2.1 Introduction 11

2.2 Model 15

2.3 Information matrix 19

2.4 Construction of optimal designs 21

2.4.1 Optimal designs for τ and γ 22

2.4.1.1 Optimal two-period designs 22

2.4.1.2 Optimal three-period designs 23

2.4.1.3 Optimal four-period designs 25

2.4.2 Optimal designs for τ in the absence of γ 28

2.5 The efficiency of two-treatment designs 30

3 Response Adaptive Repeated Measurement Designs 60

3.1 Introduction 60

3.2 Information matrix for treatment effects 64

3.3 Allocation rules for symmetric designs 73

3.4 Implementation of the rules 77

3.5 Efficiency of adaptive designs 79

3.5.1 $t = 2$ and $p = 2$ 80

3.5.2 $t = 2$ and $p = 3$ 83

3.5.3 $t = 3$ and $p = 3$ 85

3.6 Conclusion 85

4 Use of Proxy Data: An Alternative to Incomplete Data Analysis 105

4.1 Introduction 105

4.2 Model 109

4.3 Incomplete data analysis 111

4.4 Parameter estimation with proxy data 113

4.4.1 When $\mathbf{e} = \mathbf{0}$ 113

4.4.2 When $\mathbf{e} \neq \mathbf{0}$ 123

4.5 Testing procedures 129

4.6 Simulation study 130

4.7 Design considerations 133

  4.7.1 Two-period two-treatment designs 135

  4.7.2 Three-period two-treatment designs 137

4.8 Discussion 139

5 Multiple Imputation for Repeated Measures Data **153**

5.1 Introduction 153

5.2 Multiple imputation theory 156

5.3 Multiple imputation for repeated measures data 159

5.4 Simulations 164

5.5 Conclusion 167

6 Concluding Remarks **172**

6.1 Main contributions 172

6.2 Future direction 176

Bibliography **178**

List of Tables

2.1 List of all possible sequences for two-period, three-period and four-period designs 35

2.2 Values of ϕ_1, ϕ_2 and ϕ_3 at given ρ for optimal three-period designs for estimating treatment effects τ 35

2.3 Optimal three-period designs for estimating residual effects γ . 35

2.4 Values of ϕ_1 and ϕ_2 at given ρ for four-period optimal designs for estimating treatment effects τ 43

2.5 Index for two-period, three-period and four-period designs . . . 52

2.6 Efficiency of two-period designs relative to the optimal design . 53

2.7 Efficiency of three-period designs relative to the optimal design when $\rho = 0.0$ 54

2.8 Efficiency of three-period designs relative to the optimal design when $\rho = 0.3$ 54

2.9 Efficiency of three-period designs relative to the optimal design when $\rho = 0.7$ 55

2.10 Efficiency of three-period designs relative to the optimal design when ρ approaches 1.0 55

2.11 Efficiency of four-period designs relative to the optimal design when $\rho = 0.0$ 56

2.12 Efficiency of four-period designs relative to the optimal design when $\rho = 0.3$ 57

2.13 Efficiency of four-period designs relative to the optimal design when $\rho = 0.7$ 58

2.14 Efficiency of four-period designs relative to the optimal design when ρ approaches 1.0 59

4.1 Empirical sizes and powers for testing τ in Design I when $\mathbf{e} = \mathbf{0}$ 141

4.2 Empirical sizes and powers for testing γ in Design I when $\mathbf{e} = \mathbf{0}$ 142

4.3 Empirical sizes and powers for testing τ in Design IV when $\mathbf{e} = \mathbf{0}$ and $n = 10$ 143

4.4 Empirical sizes and powers for testing τ in Design IV when $\mathbf{e} = \mathbf{0}$ and $n = 20$ 144

4.5 Empirical sizes and powers for testing γ in Design IV when $\mathbf{e} = \mathbf{0}$ and $n = 10$ 145

4.6 Empirical sizes and powers for testing γ in Design IV when $\mathbf{e} = \mathbf{0}$ and $n = 20$ 146

4.7 Empirical sizes and powers for testing τ in Design I when $\mathbf{e} \neq \mathbf{0}$ and $n = 20$ 147

4.8 Empirical sizes and powers for testing γ in Design I when $\mathbf{e} \neq \mathbf{0}$ and $n = 20$ 148

4.9 Empirical sizes and powers for testing τ in Design IV when $\mathbf{e} \neq \mathbf{0}$ and $n = 20$ 149

4.10 Empirical sizes and powers for testing γ in Design IV when $\mathbf{e} \neq \mathbf{0}$ and $n = 20$ 150

4.11 Proportions of proxy data that generate estimators for τ with $c \times 100\%$ efficiency in Design II 151

4.12 Proportions of proxy data that generate estimators for τ with $c \times 100\%$ efficiency in Design IV 151

4.13 Proportions of proxy data that generate estimators for τ with $c \times 100\%$ efficiency in Design VI 152

5.1 Empirical sizes and α-adjusted powers for testing τ and γ in Design I. 169

5.2 Empirical sizes and α-adjusted powers for testing γ in Design IV. 170

5.3 Empirical sizes and α-adjusted powers for testing τ in Design IV. 171

List of Figures

2.1 Proportions of subjects in each given sequence that makes up optimal three-period designs for estimating treatment effects, when $\rho = 0.0$. 36

2.2 Proportions of subjects in each given sequence that makes up optimal three-period designs for estimating treatment effects, when $\rho = 0.3$. 37

2.3 Proportions of subjects in each given sequence that makes up optimal three-period designs for estimating treatment effects, when $\rho = 0.7$. 38

2.4 Proportions of subjects in each given sequence that makes up optimal three-period designs for estimating treatment effects, when $\rho \to 1$. 39

2.5 Proportions of subjects in each given sequence that makes up optimal three-period designs for estimating residual effects, when $\rho = 0.3$. 40

2.6 Proportions of subjects in each given sequence that makes up optimal three-period designs for estimating residual effects, when $\rho = 0.7$. 41

2.7 Proportions of subjects in each given sequence that makes up optimal three-period designs for estimating residual effects, when $\rho \to 1$. 42

2.8 Proportions of subjects in each given sequence that makes up optimal four-period designs for estimating treatment effects, when $\rho = 0.0$. 44

2.9 Proportions of subjects in each given sequence that makes up optimal four-period designs for estimating treatment effects, when $\rho = 0.3$. 45

2.10 Proportions of subjects in each given sequence that makes up optimal four-period designs for estimating treatment effects, when $\rho = 0.7$. 46

2.11 Proportions of subjects in each given sequence that makes up optimal four-period designs for estimating treatment effects, when $\rho \to 1$. 47

2.12 Proportions of subjects in each given sequence that makes up optimal four-period designs for estimating residual effects, when $\rho = 0.0$. 48

2.13 Proportions of subjects in each given sequence that makes up optimal four-period designs for estimating residual effects, when $\rho = 0.3$. 49

2.14 Proportions of subjects in each given sequence that makes up optimal four-period designs for estimating residual effects, when $\rho = 0.7$. 50

2.15 Proportions of subjects in each given sequence that makes up optimal four-period designs for estimating residual effects, when $\rho \to 1$. 51

3.1 Relative efficiency curves for adaptive two-treatment two-period designs with covariance matrix $\mathbf{V}_1$ in Section 3.5.1 with $\rho = .3$ for $M = 1$. The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., close to the line) relative to the fixed optimal design. 87

3.2 Relative efficiency curves for adaptive two-treatment two-period designs with correlation coefficient $\rho = .3$ for $M = 2$ for $P =$ 20% and $c_1 = 2$ (the solid line) and $P = 50\%$ and $c_1 = 2$ (the dashed line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$). 88

3.3 Relative efficiency curves for adaptive two-treatment two-period designs with correlation coefficient $\rho = .7$ for $M = 2$ for $P =$ 20% and $c_1 = 2$ (the solid line) and $P = 50\%$ and $c_1 = 2$ (the dashed line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$). 89

3.4 Relative efficiency curves for adaptive two-treatment two-period designs for $M = 3$. The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$). 90

3.5 Relative efficiency curves for adaptive two-treatment three-period designs when $M = 2$ and $\rho_1 = \rho_2 = \rho_3 = .3$ for $c_1 = c_2 = 2$ (the solid line) and $c_1 = 1$ and $c_2 = 2$ (the dashed line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$). 91

3.6 Relative efficiency curves for adaptive two-treatment three-period designs when $M = 2$ and $\rho_1 = \rho_3 = \sqrt{\rho_2} = .3$ for $c_1 = c_2 = 2$ (the solid line) and $c_1 = 1$ and $c_2 = 2$ (the dashed line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$). 92

3.7 Relative efficiency curves for adaptive two-treatment three-period designs when $M = 2$ and $\rho_1 = \rho_2 = \rho_3 = .7$ for $c_1 = c_2 = 2$ (the solid line) and $c_1 = 1$ and $c_2 = 2$ (the dashed line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$). 93

3.8 Relative efficiency curves for adaptive two-treatment three-period designs when $M = 2$ and $\rho_1 = \rho_3 = \sqrt{\rho_2} = .7$ for $c_1 = c_2 = 2$ (the solid line) and $c_1 = 1$ and $c_2 = 2$ (the dashed line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$). 94

3.9 Relative efficiency curves for adaptive two-treatment three-period designs when $M = 3$ and $\rho_1 = \rho_2 = \rho_3 = .3$ (the solid line), and $\rho_1 = \rho_2 = \rho_3 = .7$ (the dashed line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$). 95

3.10 Relative efficiency curves for adaptive two-treatment three-period designs when $M = 3$ and $\rho_1 = \rho_3 = \sqrt{\rho_2} = .3$ (the solid line), and $\rho_1 = \rho_3 = \sqrt{\rho_2} = .7$ (the dashed line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$). 96

3.11 Relative efficiency curves for adaptive three-treatment three-period designs when $M = 2$ and correlation coefficients $\rho_1 = \rho_3 = \sqrt{\rho_2} = .3$ and $c_1 = c_2 = 2$ using A-optimal criteria (the solid line), E-optimal criteria (the dashed line) and D-optimal criteria (the dotted line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$). 97

3.12 Relative efficiency curves for adaptive three-treatment three-period designs when $M = 2$ and correlation coefficients $\rho_1 = \rho_3 = \sqrt{\rho_2} = .3$ and $c_1 = 1, c_2 = 2$ using A-optimal criteria (the solid line), E-optimal criteria (the dashed line) and D-optimal criteria (the dotted line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$). 98

3.13 Relative efficiency curves for adaptive three-treatment three-period designs when $M = 2$ and correlation coefficients $\rho_1 = \rho_2 = \rho_3 = .3$ and $c_1 = c_2 = 2$ using A-optimal criteria (the solid line), E-optimal criteria (the dashed line) and D-optimal criteria (the dotted line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$). 99

3.14 Relative efficiency curves for adaptive three-treatment three-period designs when $M = 2$ and correlation coefficients $\rho_1 = \rho_2 = \rho_3 = .3$ and $c_1 = 1, c_2 = 2$ using A-optimal criteria (the solid line), E-optimal criteria (the dashed line) and D-optimal criteria (the dotted line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$). 100

3.15 Relative efficiency curves for adaptive three-treatment three-period designs when $M = 2$ and correlation coefficients $\rho_1 = \rho_3 = \sqrt{\rho_2} = .7$ and $c_1 = c_2 = 2$ using A-optimal criteria (the solid line), E-optimal criteria (the dashed line) and D-optimal criteria (the dotted line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$). 101

3.16 Relative efficiency curves for adaptive three-treatment three-period designs when $M = 2$ and correlation coefficients $\rho_1 = \rho_3 = \sqrt{\rho_2} = .7$ and $c_1 = 1, c_2 = 2$ using A-optimal criteria (the solid line), E-optimal criteria (the dashed line) and D-optimal criteria (the dotted line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$). 102

3.17 Relative efficiency curves for adaptive three-treatment three-period designs when $M = 2$ and correlation coefficients $\rho_1 = \rho_2 = \rho_3 = .7$ and $c_1 = c_2 = 2$ using A-optimal criteria (the solid line), E-optimal criteria (the dashed line) and D-optimal criteria (the dotted line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$). 103

3.18 Relative efficiency curves for adaptive three-treatment three-period designs when $M = 2$ and correlation coefficients $\rho_1 = \rho_2 = \rho_3 = .7$ and $c_1 = 1, c_2 = 2$ using A-optimal criteria (the solid line), E-optimal criteria (the dashed line) and D-optimal criteria (the dotted line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$). 104

Chapter 1

Introduction

1.1 Background of the thesis

This thesis focuses on the design and analysis issues that arise in the use of repeated measurement designs with a special application to clinical trials. In typical repeated measurement designs, the study subjects receive several experimental treatments over several different periods of time. The study subjects can either be treated repeatedly with the same treatments over time, or in a cross-over fashion, with different treatments over time. Responses from each of the periods for each study subject are usually correlated, and are often referred to as longitudinal or repeated measures data. To obtain efficient estimators of parameters of interest, decisions have to be made on which treatment sequences a subject should be assigned to and how many subjects should be assigned to each treatment sequence. Further, for various reasons, the resulting repeated measures data are often incomplete, so efficient analysis methods are required for repeated measures data with missing values. This thesis deals with both design and analysis issues related to repeated measurement designs.

1.1.1 Design issues

Repeated measurement designs (sometimes referred to as cross-over designs or change-over designs) are frequently used in clinical trials to compare the efficacy of several different treatments. Brown (1980) reported that in a survey of studies investigating the effects of anti-anxiety drugs on humans, 68% of the studies used the cross-over approach, which is evidence of the popularity of repeated measurement designs. Here are some of the major reasons for the widespread use of repeated measurement designs (Hedayat and Afsarinejad, 1978; Carriere and Huang, 2000 and 2001):

1. It is economical to adopt repeated measurement designs, since it can be expensive to recruit human study subjects.
2. The use of repeated measurement designs allows within-subject comparisons. This is important because the response to a treatment can vary greatly among study subjects.

3. Using repeated measurement designs can save time, especially when special training is required for the study subjects. Unlike completely randomized designs, where subjects are used only for one period, repeated measurement designs allow the repeated use of subjects who are already trained.

4. Repeated measurement designs can produce estimators for treatment effects that are more efficient than those obtained from parallel or completely randomized designs.

A simple example of a repeated measurement design is the two-treatment (A and B) two-period design with two sequences, AB and BA. In sequence AB (BA), subjects are assigned to treatment A (B) in the first period and treatment B (A) in the second period. The design is desired to produce the most efficient estimator for the treatment effects.

In constructing optimal repeated measurement designs, we first choose a model for the correlated data while considering factors involved in the experiment and the correlation structure. The factors that are usually considered are treatment effects, residual (carry-over) treatment effects (which are the effects of the treatment given in previous periods), period effects, sequence effects and subject effects (fixed or random). The most commonly used covariance structures for measurement error are:

(1) independent, in which it is assumed that the repeated measures data are uncorrelated;

(2) equicorrelated, in which it is assumed that the observations from any two different periods are equally correlated;

(3) autoregressive, in which the correlation for the observations from any two periods is an exponential function of the distance between the two periods.

Many researchers (e.g., Cheng and Wu, 1980; Hedayat and Afsarinejad, 1978; Kunert, 1983, 1985, 1991; Kushner, 1997; Laska and Meisner, 1985; Matthews, 1987) have constructed optimal designs under various model assumptions. However, these optimal designs are strongly model-dependent. For example, the optimal designs under an autoregressive (AR) error model are not the same as those obtained under an equicorrelated covariance model (Matthews, 1987). Further, clinicians may not wish to use some treatment sequences involved in the optimal designs (Carriere and Reinsel, 1992 and 1993). To alleviate these dilemmas, some researchers have considered compromised designs (e.g., Carriere and Reinsel 1992; Carriere 1994b). However, all of these investigations are based on the assumption that the covariance structures for within-subject measurements are homogeneous and known.

In this thesis, we recognize some of these problems with optimal designs, and consider a general approach to designing an experiment. First, we proceeded with constructing optimal designs in a traditional way, but used a more general model than had been considered before. We then constructed universally optimal designs by maximizing the information matrix as usual.

Next, we attempted to deal with the fact that subjects usually enter the experiment sequentially, and that one cannot know prior to the experiment which design should be chosen because of the lack of information on the correlation model. In response-adaptive randomized design, investigated by Flournoy and Rosenberger (1995), Atkinson (1982), Pocock and Simon (1975) and Efron (1971), the serial subject responses are used to update the rules for terminating the experiment. Authors such as Kushner (2000) and Cook (1995, 1996) have explored adaptive cross-over designs, where the designs are adaptively constructed as more data become available, thereby using empirical observations about the correlation pattern.

In this thesis, we also considered constructing designs adaptively. A unique improvement in our approach is that designs can be constructed for situations where the correlation structure is not only unknown but can also vary from subject to subject. The rules for allocating study subjects to one of the planned treatment sequences are then derived in such a way as to maximize the increment of the information for the treatment effects available from new subjects, based on the data from already observed subjects. Further, the allocation rules are adjusted to minimize the loss function of the increment of the information, which may be due to an incorrectly specified covariance structure.

1.1.2 Analysis issues

Now we turn to analysis issues in repeated measurement designs. In repeated measurement designs, study subjects are randomly assigned to each treatment sequence over several periods. The subjects' responses in successive periods result in a set of repeated measures data. For instance, in the two-treatment two-period design with sequences AB and BA, paired observations are obtained from the two periods for each subject in the experiment. These paired data are correlated, usually positively, since they are measured on the same subject, and are often referred to as longitudinal or repeated measures data.

More often than not, longitudinal or repeated measures data are incomplete because some subjects drop out before the experiment is complete. Missing data in repeated measurement designs can be caused by lack of treatment effects, unpleasant experiences, loss of follow-up or other unknown reasons. It is common to have missing data in experiments that take a long time to complete. Patients often fail to keep the last few follow-up appointments (Hogan and Laird, 1997; Yao et al., 1998).

Missing data can occur haphazardly or monotonically. The pattern is said to be monotonic when a subject is not observed for a certain period and all the subsequent periods. With the monotonic missing data pattern, it is possible to factorize the likelihood function as a product of conditional distributions and derive the maximum likelihood estimators (Rubin, 1987; Carriere, 1999; Patel, 1985), under the assumption that the data are missing at random.

Sometimes, especially in experiments using human subjects, missing data can occur even when the study subjects are present. For example, some patients in cancer clinical trials may simply be too sick to comply, and often, their care providers supply relevant information, called "proxy" information. Some authors have argued for the use of such proxy information (e.g., Parsons et al., 1999; Jalukar et al., 1998). However, most work on this topic has been limited to the debate between using or discarding variables with proxy information (e.g., McCallum, 1972; Wickens, 1972; Aigner, 1974; Dhrymes, 1978; Trenkler and Stahlecker, 1996). In these works and other applications, proxy information has been treated and analyzed as if it were actual.

Our aim in this thesis is to evaluate the merits of using proxy information as compared with several available approaches dealing with missing data problems in repeated measures data analyses. Three such main approaches are complete subset data only analyses, incomplete data methods (Anderson, 1957; Morrison, 1970; Patel, 1985; Carriere, 1994a and 1999) and imputation methods (Rubin, 1987; Glynn et al., 1993; Meng, 1994). Methods based only on the complete subset data have been shown to produce biased and inefficient estimators. Incomplete data methods are intended to utilize all available data, usually by using likelihood methods. Imputation methods involve making up a complete data set using various imputation strategies so that the standard statistical analysis can be applied.

Comparisons between the complete subset data analysis and the incomplete data analysis in small-sample repeated measures data have been investigated by Carriere (1994a and 1999). Her studies showed that the incomplete data methods are superior to the complete subset data method in terms of the power of testing hypotheses for treatment effects. Although the theory and applications of multiple imputation have been investigated widely since 1978, little research has been done on its strategies for small-sample repeated measures data (Carriere, 1997).

This thesis first investigated how to deal with situations where some data are missing monotonically, but proxy information is available. To our knowledge, no one has considered rigorously how to utilize proxy information in the context of trying to deal with incomplete data. The development of an appropriate method of using proxy data in small-sample repeated measures data with missing values is certain to be useful for data analysts in various disciplines.

By modeling possible bias and possible heterogeneity in the data due to some proxy values, we explored its power as compared to the incomplete data methods utilizing only the available actual data. The approach of using proxy data, in essence, corresponds to the single imputation method, which was widely considered to be inferior (Rubin, 1978, 1987) to the multiple imputation method. However, proxy information can be more informative and more personal than imputed data, as it usually comes from the next of kin. We also considered the design implications of having proxy data.

As the use of proxy is, in principle, a single imputation method, the next investigation dealt with the multiple imputation method. We first proposed a multiple imputation strategy for repeated measures data. Testing procedures for the treatment and residual effects were then proposed, building on the multiple imputation inferential technique (Rubin, 1987) and adjusting the degrees of freedom for small samples. The advantages and disadvantages of all approaches to missing, and thus deficient, data were then examined.

1.2 Thesis overview

In Chapter 2, we review all models that have been considered for constructing optimal designs. By expanding the model's capacity somewhat, we consider a general model that includes random subject effects and an autoregressive error structure for the correlated data. We show that all other models are special cases of this general model. We then construct optimal designs under this model, verifying the results against those of others for some special cases. We also consider some “nearly” optimal designs.

In Chapter 3, we construct adaptive repeated measurement designs. After obtaining the information matrix for the treatment effects, we derive the allocation rules to be used in assigning subjects to the treatment sequences, thereby minimizing the loss function of the increment of information due to an incorrectly specified covariance structure. The implications of incorporating heterogeneity among the study subjects in design construction are discussed.

In Chapter 4, we utilize proxy information for missing data, and compare this method's estimation and testing performance to that of the incomplete data method, which utilizes only the available actual data. We also consider the efficiency of designs in helping to identify situations where having proxy data is useful and where it is waste of time.

In Chapter 5, we develop a multiple imputation strategy for small-sample repeated measures data. The power of this strategy is compared to that of the incomplete data method without the use of imputation. Overall conclusions and general recommendations are given.

Finally, Chapter 6 summarizes the main contributions of this thesis to the literature. Some possible future research is suggested to expand and improve on the strategies suggested in this thesis.

Chapter 2

Optimal Two-Treatment Repeated Measurement Designs

2.1 Introduction

Repeated measurement designs have been used frequently in clinical trials to compare the efficacy of several non-curative treatments. The major appeal of repeated measurement designs lies in the fact that the estimators of direct treatment effects can be obtained efficiently, as between-subject variations can be eliminated. Although some researchers, such as Cleophas and Tavenier (1994), Brown (1980) and Cochran and Cox (1957), have questioned their feasibility, repeated measurement designs have been shown to be cost efficient as compared to completely randomized designs or parallel designs, except in a few extreme cases (Carriere and Huang, 2000 and 2001).

Generally, a repeated measurement design with t treatments and p periods can have up to t^p possible treatment sequences. Subjects may be exposed to a series of different or identical treatments. A simple two-treatment (A and B) two-period repeated measurement design randomly assigns patients to one of the two treatment sequences, AB and BA. In the sequence AB (BA), patients receive treatment A (B) in the first period and B (A) in the second period. When one sequence can be turned into another sequence by permuting the treatments, the two sequences are referred to as "dual sequences." For instance, sequence AB is the dual sequence of BA; and sequences BCA and ACB are dual sequences of ABC. We refer to a repeated measurement design as "dual balanced" if an equal number of subjects is assigned to each of the dual sequences.

Researchers have considered various models for fitting repeated measures data from cross-over design experiments. Williams (1949) introduced a model that does not include residual effects, assuming that the wash-out period will help to eliminate the treatment effects left over for the next period. For a review of Williams' model, see Matthews (1988). First-order residual effects have been added into models, concerning the persistence of treatment effects. However, Fleiss (1989) and Senn (1993) stated that using only first-order residual effects is not realistic since the treatment effects may persist beyond one period. In response to these statements, Matthews (1994a, 1994b) suggests that certain adjustments may be necessary upon a close examination of the particular situation. Another factor that determines the model and the design is the form of covariance structure for the repeated measures data, for which either an independent, equicorrelated or autoregressive structure has been employed.

Under various model assumptions, many researchers have investigated the construction of optimal repeated measurement designs (Hedayat and Afsarinejad, 1975 and 1978; Cheng and Wu, 1980; Kunert, 1983 and 1984; Laska and Meisner, 1985; Matthews, 1987; Hedayat and Zhao, 1990; Carriere and Reinsel, 1992; Kushner, 1997 and 2000). As expected, the optimal designs are highly dependent on model assumptions such as the factors included in the model and the error structure associated with the within-subject measurements. The optimal design under a certain model may no longer be optimal when one or more of its model assumptions are violated.

Typically, optimal repeated measurement designs have been constructed under the assumptions that the subjects share a common covariance matrix and the form of the error structure is known. However, in practice, the form of the covariance matrix cannot be known before the actual experiment. Some authors recommended "nearly" optimal designs that are somewhat robust to the violation of the model assumption (Carriere and Reinsel, 1992; Carriere, 1994b). A "nearly" optimal design may be robust against mild departures from the assumed form of a covariance structure, but not robust against gross departures, for example, the usual situations of an unknown covariance structure coupled with possible heterogeneity. Designs that are robust under the latter situation are discussed in Chapter 3.

In this chapter, we discuss the construction of optimal designs under a more general model to address the situation of a known covariance structure. The model to be considered here is more general than those described in the literature, in that it includes two sources of random variation—random subject error and serially correlated within-subject measurements error. Construction of optimal designs in such situations has not been considered to date. We will show how changing some parameters of the general model affects the results of the optimal designs. In this chapter, we will also evaluate the robustness of several popularly adopted two-treatment designs under this model.

The organization of this chapter is as follows. In Section 2, we describe the model to be considered. In Section 3, we obtain the information matrix for the contrast of treatment effects and residual effects. In Section 4, we construct optimal designs for two-period, three-period and four-period experiments. In Section 5, we compare and discuss the relative efficiency of several repeated measurement designs constructed by other investigators.

2.2 Model

Consider a repeated measurement design d with t treatments, p periods and s sequences. Let $\mathbf{y}_{jk} = (y_{1jk}, \ldots, y_{pjk})^T$ be the vector of observations from subject j in treatment sequence k. The model for the response $\mathbf{y}_{jk}$ is

$$\mathbf{y}_{jk} = \mathbf{X}_{jk}\boldsymbol{\beta} + \xi_{jk}\mathbf{1}_{[p]} + \boldsymbol{\epsilon}_{jk} \tag{2.2.1}$$

where $j = 1, \ldots, N_k$, $\mathbf{1}_{[p]}$ is a $p \times 1$ vector of one and N_k is the number of subjects in sequence k, $k = 1, \ldots, s$. The total number of subjects involved is $N = \sum_{k=1}^{s} N_k$. For dual-balanced designs, we have $N_k = N_{k^*}$, where k^* is the dual sequence of sequence k. The parameter $\boldsymbol{\beta} = (\mu, \boldsymbol{\pi}^T, \boldsymbol{\tau}^T, \boldsymbol{\gamma}^T)^T$ consists of the overall mean μ, the period effects $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_p)^T$, the direct treatment effect $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_t)^T$ and the first-order residual effect of the treatment given in the previous period $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_t)^T$. The design matrix $\mathbf{X}_{jk}$, which is the same for all j, $j = 1, \ldots, N_k$, is $\mathbf{X}_k = (\mathbf{X}_{1k}, \mathbf{X}_{2k})$, where $\mathbf{X}_{1k}$ relates $\mathbf{y}_{jk}$ with μ and $\boldsymbol{\pi}$, and $\mathbf{X}_{2k} = (\mathbf{X}_k^{\tau}, \mathbf{X}_k^{\gamma})$ with matrix $\mathbf{X}_k^{\tau} = (\mathbf{x}_{1k}^{\tau}, \ldots, \mathbf{x}_{pk}^{\tau})^T$ and $\mathbf{X}_k^{\gamma} = (\mathbf{x}_{1k}^{\gamma}, \ldots, \mathbf{x}_{pk}^{\gamma})^T$ relating $\mathbf{y}_{jk}$ with $\boldsymbol{\tau}$ and $\boldsymbol{\gamma}$, respectively. The component $\mathbf{x}_{ik}^{\tau}$ is a vector composed of a 1 and $p-1$ 0's. The position of 1 is an index of the treatment assigned in period i to sequence k. For instance, in a three-treatment (A, B and C) design, the vector $\mathbf{x}_{ik}^{\tau} = (0, 0, 1)^T$ indicates that treatment C is assigned to subjects in sequence k in period i. Further, we have $\mathbf{x}_{ik}^{\gamma} = \mathbf{x}_{i-1,k}^{\tau}$

for $i = 2, \ldots, p$. Since there is no residual effect in the first period, we have $\mathbf{x}^{\gamma}_{1k} = \mathbf{0}$. In the literature, the subject effect ξ_{jk} has been treated as either fixed or random. Random subject effects are typically assumed to have a normal distribution with mean 0 and variance σ^2_ξ, independent of the random error $\boldsymbol{\epsilon}_{jk}$. The random error $\boldsymbol{\epsilon}_{jk} = (\epsilon_{1jk}, \ldots, \epsilon_{pjk})^T$ is assumed to have a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_{jk}$. The covariance matrix of $\mathbf{y}_{jk}$ is denoted as $\mathbf{V}_{jk}$.

In this chapter, we consider model (2.2.1) for two-treatment (A and B) design with a full rank design matrix. The parameter now is $\boldsymbol{\beta} = (\mu, \boldsymbol{\pi}^T, \tau, \gamma)^T$. We are interested in estimating and testing for the contrast of direct treatment effects $\tau = (\tau_A - \tau_B)/2$ and the contrast of residual treatment effects $\gamma = (\gamma_A - \gamma_B)/2$. The component of the design matrix, x^{τ}_{ik}, takes a value of 1, if a subject in sequence k takes treatment A in period i. Otherwise, this component has a value of -1. The design matrix $\mathbf{X}_{2k} = (\mathbf{x}^{\tau}_k, \mathbf{x}^{\gamma}_k)$, with vectors $\mathbf{x}^{\tau}_k = (x^{\tau}_{1k}, \ldots, x^{\tau}_{pk})^T$ and $\mathbf{x}^{\gamma}_k = (x^{\gamma}_{1k}, \ldots, x^{\gamma}_{pk})^T$. We further assume that all subjects are associated with the same covariance matrix, i.e., $\boldsymbol{\Sigma}_{jk} = \boldsymbol{\Sigma}$. Thus, the covariance matrix of $\mathbf{y}_{jk}$ is $\mathbf{V}_{jk} = \mathbf{V}$, for $j = 1, \ldots, N_k$ and $k = 1, \ldots, s$.

Several variations of the above full rank model have been considered for constructing optimal designs:

(a) fixed subject effects and an independent error assumption ($\mathbf{V} = \boldsymbol{\Sigma} = \sigma^2_\varepsilon \mathbf{I}_{[\mathbf{p}]}$, $\sigma^2_\xi = 0$)

(b) random subject effects and an independent error assumption ($\mathbf{V} = \sigma_\varepsilon^2 \mathbf{I}_{[\mathbf{p}]} + \sigma_\xi^2 \mathbf{1}_{[\mathbf{p}]} \mathbf{1}_{[\mathbf{p}]}^{\mathbf{T}}$)

(c) fixed subject effects and an autoregressive error assumption ($(\mathbf{V})_{rc} = \sigma_\varepsilon^2 \phi^{|r-c|}/(1-\phi^2)$, where $-1 < \phi < 1$, $(\mathbf{V})_{rc}$ is the element of $\mathbf{V}$ on the r^{th} row and c^{th} column)

Model (a) has been studied by Hedayat and Afsarinejad (1978), Cheng and Wu (1980), Laska, Meisner and Kushner (1983), Kunert (1983, 1984) and Hedayat and Zhao (1990). Model (b) was introduced to address the dependence among ϵ_{ijk}, $i = 1, \ldots, p$, since $\{y_{ijk}\}$ are measured on the same subject. The within-subject correlation is $\rho = \sigma_\xi^2/\sigma^2$, where $\sigma^2 = \sigma_\xi^2 + \sigma_\epsilon^2$. Laska and Meisner (1985) and Carriere and Reinsel (1992 and 1993) have investigated optimal designs for model (b). Note that variance-component models (a) and (b) assume *a priori* that the correlation is positive. Models (a) and (b) are related, in that estimators from a fixed subject effects model (a) can be obtained using the estimators from a random subjects effect model (b) by letting $\rho \to 1$, as shown by Carriere and Reinsel (1992). Considering that the correlation among within-subject measurement is decreasing with time exponentially, Matthews (1987), Kunert (1991) and Kushner (1997) studied optimal designs under model (c). Kunert (1985) considered a similar model without residual treatment effects, while Laska and Meisner (1985) discussed model (c) without explicit subject effects.

Since repeated or longitudinal data are likely to exhibit many qualitatively different sources of random variations (Diggle, 1988), we consider another variant of the model. We consider a model with two sources of such random variations—random subject effects and serially correlated errors in the form of a first-order autoregression (AR(1)), i.e.,

$$\epsilon_{ijk} = \phi\epsilon_{i-1,jk} + \eta_{ijk} \tag{2.2.2}$$

for $i > 1$, and $\epsilon_{1jk} = \eta_{1jk}$, where $\eta_{1jk} \sim N(0, \sigma_\varepsilon^2/(1-\phi^2))$ and $\eta_{ijk} \sim N(0, \sigma_\varepsilon^2)$ for $i > 1$. Under this model, the elements in the covariance matrix $\mathbf{V}$ of $\mathbf{y}_{jk}$ are

$$(\mathbf{V})_{rc} = \sigma_\varepsilon^2 \phi^{|r-c|}/(1-\phi^2) + \sigma_\xi^2 \tag{2.2.3}$$

where, $-1 < \phi < 1$. We label such a model with random subject effects and AR(1) error structure as model (d), namely:

(d) random subject effects and an autoregressive error assumption for $\mathbf{V}$ as in (2.2.3)

2.3 Information matrix

The information matrix for the treatment contrast and the residual treatment contrast (τ, γ) in a two-treatment design under our model (d) is (Carriere and Huang, 2000)

$$\mathbf{I}(\tau, \gamma) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}, \tag{2.3.1}$$

where

$$I_{11} = \sum_k N_k[p + \phi^2(p-2) - 2\phi q_k - b(l_k - \phi l_k^{**})^2]/\sigma_\varepsilon^2,$$

$$I_{12} = \sum_k N_k[q_k + \phi^2 \tilde{q}_k - \phi(f_k + p - 1) - b(l_k - \phi l_k^{**})(\tilde{l}_k - \phi l_k^{*})]/\sigma_\varepsilon^2,$$

and

$$I_{22} = \sum_k N_k[p - 1 + \phi^2(p-2) - 2\phi \tilde{q}_k - b(\tilde{l}_k - \phi l_k^{*})^2]/\sigma_\varepsilon^2,$$

for $b = (1-\phi)^2\rho/\{(1-\rho) + (1-\phi)[p - (p-2)\phi]\rho\}$, $q_k = \sum_{i=1}^{p-1} x_{ik}^{\tau} x_{i+1,k}^{\tau}$, $l_k = \sum_{i=1}^{p} x_{ik}^{\tau}$, $\tilde{l}_k$ is the same as l_k but summing up to $p-1$, l_k^{*} and $\tilde{q}_k$ are the same as l_k and q_k but summing up to $p-2$, l_k^{**} is the same as l_k but summing from 2 up to $p-1$, and $f_k = \sum_{i=1}^{p-2} x_{ik}^{\tau} x_{i+2,k}^{\tau}$.

Remark 2.3.1 *When* $p = 2$, $l_k^{*} = l_k^{**} = f_k = \tilde{q}_k = 0$.

Remark 2.3.2 *When* ϕ *is zero, the within-subject measurement error is not serially correlated, but exchangeable. In this case, the information matrix corresponds to that under model (b), and the information matrix (2.3.1) reduces* *to*

$$\mathbf{I}(\tau,\gamma) = \frac{\sum N_k}{\sigma_\epsilon^2}\begin{bmatrix} p - b_0 l_k^2 & q_k - b_0 l_k \tilde{l}_k \\ q_k - b_0 l_k \tilde{l}_k & p - 1 - b_0 \tilde{l}_k^2 \end{bmatrix}$$

as shown in Carriere (1994), where $b_0 = \rho/[1 + (p-1)\rho]$.

The connection between model (c) and (d) is illustrated by the following theorem.

Theorem 2.3.1 *If we let* $\rho \to 1$, *the information matrix (2.3.1) becomes*

$$\mathbf{I}(\tau,\gamma) = \frac{4}{\sigma_\epsilon^2}\begin{bmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{bmatrix}$$

as shown in Matthew (1987) under the fixed subject effect model. The components are $h_{11} = 2\phi\sum_k N_k c_k + 2R(1-\phi)z_1$, $h_{22} = 2\phi\sum_k N_k \tilde{c}_k + 2R(1-\phi)z_2 + z_4$, *and* $h_{12} = R(1-\phi)z_3 - \sum_k N_k(c_k - \phi e_k + \phi^2 \tilde{c}_k)$, *where* $c_k = \sum_{i=1}^{p-1} \chi_i^k$ *with* $\chi_i^k = 0$ *if the same treatment is administered in the two successive periods, and 1 otherwise. Here,* $\tilde{c}_k$ *is the same as* c_k *but summing up to* $p-2$, $e_k = c_k + \tilde{c}_k - \chi_1^k - 2\sum_{i=1}^{p-2} \chi_i^k \chi_{i+1}^k$, $z_1 = \sum_k N_k(r_k - \phi r_k^*)(r_{k\cdot} - \phi r_{k\cdot}^*)$, $z_2 = \sum_k N_k(\tilde{r}_k - \phi\bar{r}_k)(\tilde{r}_{k\cdot} - \phi\bar{r}_{k\cdot})$, $z_3 = \sum_k N_k[(r_k - \phi r_k^*)(\tilde{r}_{k\cdot} - \phi\bar{r}_{k\cdot}) + (r_{k\cdot} - \phi r_{k\cdot}^*)(\tilde{r}_k - \phi\bar{r}_k)]$, $z_4 = s[p-1-\phi(p-3)]/4[p-\phi(p-2)]$, *and* $R = 1/[p-(p-2)\phi]$. *Note that* r_k *is the number of occurrences of treatment A on sequence k in all periods,* r_k^* *is that for periods 2 to* $p-1$, $\tilde{r}_k$ *is that for the first* $p-1$ *periods, and finally,* $\bar{r}_k$ *is* *that for the first $p-2$ periods, as defined in Matthews (1987). All summations are over one-half of all sequences.*

PROOF. When $\rho \to 1$, we have that $q_k = p - 1 - 2c_k$, $\tilde{q}_k = p - 2 - 2\tilde{c}_k$, $l_k = 2r_k - p$, $l_k^{**} = 2r_k^{*} - (p-2)$, $l_k^{*} = 2\tilde{r}_k - (p-1)$, $\bar{l}_k = 2\bar{r}_k - (p-2)$, $f_k = (p-2) - 2e_k$, and the theorem follows upon algebraic manipulations.

□

2.4 Construction of optimal designs

For two-treatment designs, the universal optimal design is the one that maximizes the information for τ or γ, which is obtainable from the information matrix provided in equation (2.3.1). The optimal design depends on the level of autoregressive coefficient ϕ and the ratio ρ between the two variances σ_ξ^2 and $\sigma^2 = \sigma_\xi^2 + \sigma_\varepsilon^2$.

For two-period designs, we obtain the optimal designs for estimating τ and γ by maximizing their information directly. For dual-balanced three-period and four-period designs, we used the software MAPLE to obtain the optimal designs. Here, we outline the steps taken to construct the optimal designs. First, we found the sequences in the optimal designs for selected values of ρ using Lemma 3.2 (Kushner, 1997). Second, the proportions of subjects to each sequence in the optimal design were obtained by applying equation 6.4 (Kushner, 1997). Third, the information on treatment effects for the optimal designs was obtained and the values of ϕ were determined. Table 2.1 provides a list of all possible sequences in two-period, three-period and four-period designs, and the number of subjects for each of the sequences. The sequences shown as a pair in Table 2.1 are dual sequences to each other. An optimal design allocates an equal number of subjects to a sequence and its dual sequence, as Kushner (1997) found under a model with fixed subject effects.

2.4.1 Optimal designs for τ and γ

2.4.1.1 Optimal two-period designs

For two-treatment two-period designs, there are four possible sequences to consider: AA, AB and their duals, as listed in Table 2.1.

The optimal two-period design for estimating τ is the design with an equal number of subjects assigned to the sequences AA, AB and their duals, regardlessly of the values of ρ and ϕ. This result coincides with those under no serial correlation model (Kershner and Federer, 1981).

For estimating γ, the optimal design assigns $N_1 = (1+\phi)(1+\rho-2\rho\phi)N/[4(1-\rho\phi^2)]$ subjects to sequence AA and its dual, and $N/2 - N_1$ to sequence AB and its dual. Note that the optimal design under no serial correlation model is to use sequence AA and its dual (Carriere and Reinsel, 1992).

2.4.1.2 Optimal three-period designs

For two-treatment three-period designs, there are eight possible sequences to consider: AAA, AAB, ABA, ABB and their duals, as listed in Table 2.1. Following is a list of optimal three-period designs for treatment effects τ with respect to the values of ρ and ϕ.

When $.3 \leq \rho < 1$, we have five different optimal designs determined by ϕ:

(a) $-1 < \phi < 0$, $N_2 = \frac{\phi(5\rho\phi-3\rho-\phi-1)}{4[\phi^2(4\rho-1)-4\rho\phi+\rho]}N$, $N_4 = \frac{N}{2} - N_2$

(b) $0 < \phi < \phi_1$, $N_3 = \frac{\phi[\phi^3(5\rho^2-\rho)+\phi^2(-9\rho^2+\rho)+\phi(7\rho^2-3\rho)-2\rho^2+\rho+1]}{4[\phi^2\rho+\phi(3\rho-1)-2\rho-1]^2}N$, $N_4 = \frac{N}{2} - N_3$

(c) $\phi_1 < \phi < \phi_2$, $N_2 = \frac{\phi^4(15\rho^2-8\rho+1)+\phi^3(-6\rho^2-\rho+1)+\phi^2(-19\rho^2+1)+\phi(17\rho^2+5\rho)-4\rho^2-2\rho}{4[\phi^4(4\rho^2-\rho)+\phi^3(-5\rho+1)+\phi^2(-7\rho^2+5\rho)+\phi(5\rho^2-\rho)-\rho^2]}N$,
$N_4 = \frac{N}{2} - N_2$

(d) $\phi_2 < \phi < \phi_3$, $N_2 = \frac{N}{2}$

(e) $\phi_3 < \phi < 1$, $N_2 = \frac{\phi^4 g_1+\phi^3 g_2+\phi^2 g_3+\phi g_4-4\rho^2+2\rho+2+x[\phi^2(5\rho-1)+\phi(4\rho-2)-4\rho-2]}{8x[\phi^2\rho+\phi(\rho-1)-\rho]}N$,
$N_3 = \frac{N}{2} - N_2$, where $g_1 = 11\rho^2-8\rho+1$, $g_2 = -5\rho^2+4\rho+1$, $g_3 = -13\rho^2+2\rho-1$, $g_4 = 14\rho^2-6\rho$, $g_5 = 17\rho^2-10\rho+1$, $g_6 = -14\rho^2+4\rho+2$, $g_7 = 3\rho^2-6\rho-1$, $g_8 = -2\rho^2+2$ and $x = \sqrt{\phi^4 g_5 + \phi^3 g_6 + \phi^2 g_7 + \phi g_8 + \rho^2 + 2\rho + 1}$.

The information for τ for the five optimal designs above is denoted as $I_{(a)}$, $I_{(b)}$, $I_{(c)}$, $I_{(d)}$ and $I_{(e)}$. The ϕ_1, ϕ_2, ϕ_3 are the threshold values that provide the solutions for $I_{(b)} = I_{(c)}$, $I_{(c)} = I_{(d)}$ and $I_{(d)} = I_{(e)}$, respectively. Table 2.2 illustrates ϕ_1, ϕ_2 and ϕ_3 for selected values of ρ.

For small ρ, we reported only the cases of $\rho = 0$, $\rho = 0.1$ and $\rho = 0.2$.

When $\rho = 0.2$, the optimal designs for τ are:

(f) $-1 < \phi < 0$, $N_2 = \frac{2\phi}{\phi^2+4\phi-1}N$, $N_4 = \frac{N}{2} - N_2$

(g) $0 < \phi < \phi_1$, $N_3 = \frac{-\phi(\phi^2+2\phi-7)}{(\phi^2-2\phi-7)^2}N$, $N_4 = \frac{N}{2} - N_3$

(h) $\phi_1 < \phi < 1$, $N_2 = \frac{\phi^4-10\phi^3+7\phi^2+4\phi-14+(3\phi+7)\sqrt{-2\phi^4+14\phi^3-13\phi^2+12\phi+9}}{-4(\phi^2-4\phi-1)\sqrt{-2\phi^4+14\phi^3-13\phi^2+12\phi+9}}N$, $N_3 =$ $\frac{N}{2} - N_2$

When $\rho = 0.1$, the optimal designs for τ are:

(i) $-1 < \phi < -.2857$, $N_2 = \frac{9\phi^4-63\phi^3+283\phi^2-368\phi+220}{4(\phi^2-11\phi+10)^2}N$, $N_1 = \frac{N}{2} - N_2$

(j) $-.2857 < \phi < 0$, $N_2 = \frac{\phi(5\phi+13)}{4(6\phi^2+4\phi-1)}N$, $N_4 = \frac{N}{2} - N_2$

(k) $0 < \phi < \phi_1$, $N_3 = \frac{\phi(-5\phi^3+\phi^2-23\phi+108)}{4(\phi^2-7\phi-12)}N$, $N_4 = \frac{N}{2} - N_3$

(l) $\phi_1 < \phi < 1$, $N_2 = \frac{31\phi^4+135\phi^3-93\phi^2-46\phi+216+(-5\phi^2-16\phi-24)\sqrt{17\phi^4+226\phi^3-157\phi^2+198\phi+121}}{8(\phi^2-9\phi-1)\sqrt{17\phi^4+226\phi^3-157\phi^2+198\phi+121}}N$, $N_3 = \frac{N}{2} - N_2$

When $\rho = 0$ and thus $\sigma_\xi^2 = 0$, the optimal designs are:

(m) $-1 < \phi < 0$, $N_1 = \frac{\phi^4-\phi^3-\phi^2+2-(\phi^2+2\phi+2)\sqrt{\phi^4-2\phi^3-\phi^2-2\phi+1}}{-8\phi\sqrt{\phi^4-2\phi^3-\phi^2-2\phi+1}}$, $N_4 = \frac{N}{2} - N_1$

(n) $0 < \phi < 1$, $N_2 = \frac{\phi^4+\phi^3-\phi^2+2-(\phi^2+2\phi+2)\sqrt{\phi^4+2\phi^3-\phi^2+2\phi+1}}{-8\phi\sqrt{\phi^4+2\phi^3-\phi^2+2\phi+1}}$, $N_3 = \frac{N}{2} - N_2$

For the model without subject effects ($\sigma_\xi^2 = 0$), the sequences in our optimal designs are in accordance with the findings of Kushner (1997). However, Kushner (1997) did not provide the proportions of each sequence required for the optimal design.

Our investigation appears to indicate that the optimal designs for τ generally either use the sequences AAB, ABB and their duals, or assign most of the subjects (> 90%) to either AAB and its dual or ABB and its dual. For moderate to large values of ϕ, a moderately high level of ρ requires the use of the sequence AAB and its dual.

Figures 2.1-2.4 illustrate how the proportions N_1/N, N_2/N, N_3/N and N_4/N change with the autocorrelation ϕ and the variance ratio ρ to produce the three-period optimal designs for estimating τ when $\rho = 0$, .3, .7 and when ρ approaches 1.

The optimal three-period designs for estimating residual effects γ, constructed in a similar way, are summarized in Table 2.3 and Figures 2.5-2.7.

2.4.1.3 Optimal four-period designs

For two-treatment four-period designs, there are sixteen sequences to consider: AAAA, AAAB, AABA, AABB, ABBB, ABBA, ABAB, ABAA and their duals (see Table 2.1). We now document the optimal four-period designs for estimating treatment effects.

When $.3 \le \rho < 1$, the optimal design allocates N_4 of the subjects to sequences AABB and its dual, and $N/2 - N_4$ to ABBA and its dual for $-1 <$

$\phi < \phi_2$ and $\phi \neq 0$; $N/2$ subjects to ABBA and its dual for $\phi_2 < \phi < 1$, where

$$N_4 = \frac{\rho^2 f_1 + \rho f_2 - 2\phi^2 - 3 + (\rho f_3 + 2\phi^2 + 3)g_1}{8\phi\rho(\phi^3 - 3\phi^2 + 3\phi - 1)g_1} N \qquad (2.4.1)$$

with $f_1 = 4\phi^7 - 28\phi^6 + 77\phi^5 - 120\phi^4 + 137\phi^3 - 149\phi^2 + 98\phi - 24$, $f_2 = 4\phi^5 - 16\phi^4 + 29\phi^3 - 25\phi^2 + 35\phi - 17$, $f_3 = 4\phi^4 - 12\phi^3 + 11\phi^2 - 16\phi + 8$ and $g_1 = \sqrt{\rho^2 f_4 + \rho f_5 + 1}$, $f_4 = 4\phi^8 - 28\phi^7 + 89\phi^6 - 156\phi^5 + 166\phi^4 - 126\phi^3 + 85\phi^2 - 42\phi + 9$ and $f_5 = -4\phi^5 + 12\phi^4 - 14\phi^3 + 12\phi^2 - 14\phi + 6$.

For small ρ (i.e., $0, .1, .2$), the optimal design allocates $N/2$ subjects to AABB and its dual for $-1 < \phi < \phi_1$; N_4 of the subjects in equation (2.4.1) to sequences AABB and its dual, and $N/2 - N_4$ to ABBA and its dual for $\phi_1 < \phi < \phi_2$ and $\phi \neq 0$; and $N/2$ subjects to ABBA and its dual for $\phi_2 < \phi < 1$.

In Table 2.4, the values of ϕ_1 and ϕ_2 that determine four-period optimal designs are listed for selected values of ρ. The four-period optimal designs are summarized in Figures 2.8-2.11.

For residual effects, when $.1 \leq \rho < 1$, the optimal designs allocate N_6 of the subjects to *ABBA* and its dual, $N/2 - N_6$ to ABBB and its dual for negative ϕ, where N_6 is

$$N_6 = \frac{\rho(4\phi^3 - 14\phi^2 + 9\phi - 1) + 3\phi - 1}{4[\rho(2\phi^3 - 7\phi^2 + 7\phi - 2) + \phi - 1]} N.$$

It allocates N_8 of the subjects to $ABAA$ and its dual, $N/2 - N_8$ to AABB and its dual for positive ϕ, where N_8 is obtained from

$$N_8 = \frac{\rho^2 f_6 + \rho f_7 + \phi^4 + \phi^2 + 2\phi + 2 + (\rho f_8 - \phi^2 + \phi - 2)g_2}{4[\rho(2\phi^3 - 7\phi^2 + 5\phi - 1) + \phi]g_2} N,$$

with $f_6 = 4\phi^8 - 24\phi^7 + 54\phi^6 - 62\phi^5 + 36\phi^4 - 10\phi^3 + 53\phi^2 - 66\phi + 21$, $f_7 = 4\phi^6 - 12\phi^5 + 11\phi^4 - 6\phi^3 - 8\phi^2 - 14\phi + 13$, $f_8 = -2\phi^4 + 8\phi^3 - 13\phi^2 + 15\phi - 6$, $g_2 = \sqrt{\rho^2 f_9 + \rho f_{10} + \phi^4 + 2\phi^3 - \phi^2 + 2\phi + 1}$, with $f_9 = 4\phi^8 - 24\phi^7 + 52\phi^6 - 44\phi^5 + 80\phi^3 - 27\phi^2 - 32\phi + 16$ and $f_{10} = 4\phi^6 - 8\phi^5 - 8\phi^4 + 22\phi^3 - 30\phi^2 + 2\phi + 8$.

When there are random subject effects, ABBA and AABB and their duals are needed for estimating γ just as in the case of estimating τ, but ABBB and ABAA and their duals are also important, depending on the sign of ϕ.

For $\rho = 0$, the optimal design is different from the one mentioned previously for negative ϕ, which allocates N_2 of the subjects to the sequence AAAB and its dual, and remaining subjects to the sequence ABBA and its dual, where N_2 is

$$N_2 = \frac{\phi^4 + \phi^2 - 2\phi + 2 - (\phi^2 + \phi + 2)\sqrt{\phi^4 - 2\phi^3 - \phi^2 - 2\phi + 1}}{-4\phi\sqrt{\phi^4 - 2\phi^3 - \phi^2 - 2\phi + 1}} N.$$

Figures 2.12-2.15 summarize these four-period optimal designs for γ when $\rho = 0$, .3, .7, and when ρ approaches 1.

For $\rho = 0$, the optimal four-period designs we obtained include the same sequences as those reported by Kushner (1997). The optimal three-period and four-period designs we obtained for $\rho \to 1$ are in accordance with those of Matthews (1987).

2.4.2 Optimal designs for τ in the absence of γ

When there are no residual effects in the model, the optimal designs are different from those obtained above.

Lemma 2.4.1 *Under our model (d) without residual effects, the information matrix (2.3.1) for the treatment effects is*

$$I_{11} = \sum_k N_k[p + \phi^2(p-2) - 2\phi q_k - b(l_k - \phi l_k^{**})^2]/\sigma_\varepsilon^2. \qquad (2.4.2)$$

Theorem 2.4.1 *If there are no residual effects and $\phi > 0$, the optimal design for the contrast of treatment effects τ is to alternate the two different treatments.*

PROOF. The optimal design is obtained by maximizing 2.4.2 in Lemma 2.4.1. Note that $q_k \geq -(p-1)$ with the equality attained for a sequence that alternates the treatments. Also, $(l_k - \phi l_k^{**})^2$ is minimized when the sequence balances the numbers of treatments as nearly as possible in all periods, and in the middle $p-2$ periods as well. For even p, both q_k and $(l_k - \phi l_k^{**})^2$ are minimized when the treatments are alternated in sequence k. For odd p, the choice is between the sequences with alternating treatments in p periods and the sequences alternating treatments in the first $p-1$ periods. The information reduction of the former design to the latter is $4\phi(b-1) < 0$, leading to Theorem 2.4.1, that the optimal design is to alternate treatments in each period of the sequences.

$\square$

If $\phi < 0$ and there are no residual effects, the optimal designs depend on the values of ϕ. Here, we present three examples for optimal designs with negative ϕ for two-period, three-period and four-period designs.

Example 2.4.1 *For two-period two-treatment designs, the optimal design for the treatment effect when* $\gamma = 0$ *and* $\phi < 0$ *is* AB *and its dual when* $\rho > \phi/(\phi^2 + \phi - 1)$; AA *and its dual when* $\rho < \phi/(\phi^2 + \phi - 1)$.

Example 2.4.2 *For three-period two-treatment designs, the optimal design for the treatment effect when* $\gamma = 0$ *and* $\phi < 0$ *is* AAB, ABB *and their duals when* $\rho > \phi/(3\phi - 2)$; AAA *and its dual when* $\rho < \phi/(3\phi - 2)$.

Example 2.4.3 *For four-period two-treatment designs, the optimal design for the treatment effect when* $\gamma = 0$ *and* $\phi < 0$ *is* $AAAA$ *and its dual when* $\rho < -\phi/(-5\phi - 2\phi^2 + 2\phi^3 + 4)$; *and* $AABB$ *and its dual when* $\rho > -\phi/(-5\phi - 2\phi^2 + 2\phi^3 + 4)$.

2.5 The efficiency of two-treatment designs

We now compare the efficiency of some commonly used designs with that of optimal designs. For estimating the treatment effect contrast τ and residual effect contrast γ, the efficiency of the selected repeated measurement designs is computed as the ratio of the information for τ or γ of the selected design to that of the optimal design. The designs that have been frequently discussed and employed are listed in Table 2.5.

For two-period designs, Design I is the optimal design under no residual effects model (Grizzle, 1965; Laska and Meisner, 1985). Design II has been shown to be optimal under various models (Carriere and Reinsel, 1992; Laska and Meisner, 1985) and also in Section 2.4.1 of this thesis. Design III is the parallel group two-period two-treatment design.

Design IV was shown to be the universally optimal design under the equicorrelated covariance model (Kershner and Federer, 1981; Laska et al., 1983). Under an AR(1) error model with fixed or random subject effects, the optimal design depends on the values of ϕ, the autoregressive coefficient, as well as ρ. For positive ϕ with fixed subject effects, it has been shown that Design VII remains "nearly" optimal among three-period designs except when the values of ϕ are close to 1 or -1 (Matthews, 1987). When ϕ is very high and positive ($> .4$), the universally optimal design allocates all subjects to AAB and BBA sequences. On the other hand, when ϕ is moderately negative, the optimal design allocates subjects to AAB and ABB and their duals, with a higher allocation of subjects to ABB and its dual. The optimal design under the no subject effect model with an AR(1) error is quite different; it is one with AAB, ABB and their duals, with over 90% of the subjects allocated to the sequence AAB and its dual (Laska et al., 1983). A compromise between statistical efficiency and ethical considerations led Carriere (1994) to recommend Design VI as the "nearly" optimal three-period design. It performed very competitively under model (b) with various within-subject correlation coefficients and various forms of residual effects. Carriere (1994) discussed allocating higher proportions of subjects to the ABB and BAA sequences than to the AAB and BBA sequences. Such a design corresponds to the optimal designs constructed when ϕ is slightly negative under the AR(1) error and fixed subject effects model (Matthews, 1987). Design VIII is the parallel group three-period two-treatment design. Kershner (1986) considered Designs V and VIII under the second-order residual effects model.

For four-period designs, Design XIV is considered to be the optimal design under an equicorrelated error model with random subject effects (Laska and Meisner, 1985). Design IX has been suggested as the optimal design for small negative ϕ under an AR(1) error model without subject effects (Laska and Meisner, 1985). Design X was shown to be superior to other designs under an AR(1) error model with fixed subject effects for relatively large ϕ (Matthews, 1987). Designs XI, XII, XIII, XV and XVI have also been considered by Matthews (1987).

Tables 2.6-2.14 tabulate the efficiencies of Designs I-XVI for various values of ϕ and ρ. The case of $\rho = 0$ represents the model without subject effects, as discussed by Laska and Meisner (1985), Matthews (1987) and Kushner (1997). The case of $\rho \to 1$ corresponds to the model with fixed subject effects. Two other values (i.e., .3 and .7) were chosen to study the efficiency of the designs. The efficiencies obtained for Designs IV, VI, VII, IX-XVI when $\rho \to 1$ are in accordance with the findings of Matthews (1987).

For two-period designs, Design II is the optimal design with an equal allocation of subjects in the four sequences. Designs I and III perform identically and not well for estimating τ, with efficiency lower than 50%. For residual effects, Design II is competitive for moderate value of ϕ when repeated measures are slightly correlated. When the within-subject correlation is relatively large, Designs II and III perform reasonably well for negative and positive ϕ, respectively. Under the model with fixed subject effects, Design III is optimal for γ, which is also true under a model with independent random errors and random subject effects (Laska and Meisner, 1985).

Among all three-period designs considered, we found that Design VI with an equal allocation to each sequence is extremely competitive as compared with the optimal designs for treatment effects we obtained in the previous section; the efficiency was higher than 93% for all combinations of ρ and ϕ. Design IV is also quite competitive for moderate values of ϕ, i.e., $-.4 \leq \phi \leq .6$. For residual effects, Design IV performs well for the case $\phi \geq -.4$ when within-subject correlation is larger than .3, and is recommended for situations with positive ϕ and no subject effects (i.e., $\rho = 0$).

For four-period designs, Design IX has a high efficiency ($\geq$ 90%) for non-positive ϕ for treatment effects, as does Design X for non-negative ϕ. Design XIV performs reasonably well for all moderate values of ϕ; the efficiency is 84% or higher. When there are no subject effects in the model, Design XV and XVI are also very good, provided $-.4 \leq \phi \leq .4$. For residual effects, Design XV is nearly optimal for non-negative ϕ and all values of ρ; the efficiency is higher than 99%. Designs X and XVI are also excellent for estimating residual effects for non-positive ϕ when ρ is not extremely small. Design VII, recommended by Ebbutt (1984), does not perform very well.

The "best" design would be a robust design with high efficiency as compared with the optimal designs for various values of ϕ and ρ. For two-period designs, Design II is optimal for τ. For residual effects, although none of them seems to be robust, Design II achieves reasonably high efficiency for low to moderate values of ρ. It appears that Design VI and Design IV achieve the most robustness for treatment effects and residual effects, respectively, in a three-period experiment. For four-period experiments, Design XIV appears to be the most robust one for both treatment effects and residual effects. For various ρ and ϕ, Design XIV attained a minimum efficiency of 75%.

Table 2.1: List of all possible sequences for two-period, three-period and four-period designs

$p = 2$		$p = 3$		$p = 4$	
Sequences	(N_k, N_{k^*})	Sequences	(N_k, N_{k^*})	Sequences	(N_k, N_{k^*})
(AA, BB)	(N_1, N_1)	(AAA, BBB)	(N_1, N_1)	(AAAA, BBBB)	(N_1, N_1)
(AB, BA)	(N_2, N_2)	(AAB, BBA)	(N_2, N_2)	(AAAB, BBBA)	(N_2, N_2)
		(ABA, BAB)	(N_3, N_3)	(AABA, BBAB)	(N_3, N_3)
		(ABB, BAA)	(N_4, N_4)	(AABB, BBAA)	(N_4, N_4)
				(ABBB, BAAA)	(N_5, N_5)
				(ABBA, BAAB)	(N_6, N_6)
				(ABAB, BABA)	(N_7, N_7)
				(ABAA, BABB)	(N_8, N_8)

Table 2.2: Values of ϕ_1, ϕ_2 and ϕ_3 at given ρ for optimal three-period designs for estimating treatment effects τ

ρ	ϕ_1	ϕ_2	ϕ_3
0.1	.3177		
0.2	.3634		
0.3	.3900	.4079	.4511
0.4	.4086	.4412	.5076
0.5	.4226	.4689	.5484
0.6	.4339	.4906	.5806
0.7	.4433	.5107	.6076
0.8	.4512	.5287	.6316
0.9	.4581	.5455	.6537
$\to 1.0$	.4641	.5616	.6750

Table 2.3: Optimal three-period designs for estimating residual effects γ

	ϕ			
ρ	$(-1, \frac{2\rho}{3\rho-1})$	$(\frac{2\rho}{3\rho-1}, \frac{1-\rho-\sqrt{1-2\rho+5\rho^2}}{2\rho})$	$(\frac{1-\rho-\sqrt{1-2\rho+5\rho^2}}{2\rho}, 0)$	$(0, 1)$
0	AAB$(\frac{1}{2})$			ABB$(\frac{1}{2})$
$(0, .2)$	AAB$(\frac{1}{2})$	AAB$(\frac{(1-\rho)(\phi-2)}{4(\rho\phi-1)})$ AAA$(\frac{3\rho\phi-2\rho-\phi}{4(\rho\phi-1)})$	ABA $(\frac{\phi[1+\phi+\rho(3-5\phi)]}{4[\rho(\phi^2+3\phi-2)-\phi-1]})$ ABB $(\frac{(\phi+1)[\rho(7\phi-4)-(\phi+2)]}{4[\rho(\phi^2+3\phi-2)-\phi-1]})$	ABB$(\frac{1}{2})$
$[.2, 1)$	AAB$(\frac{(1-\rho)(\phi-2)}{4(\rho\phi-1)})$ AAA$(\frac{3\rho\phi-2\rho-\phi}{4(\rho\phi-1)})$			
$\to 1$	ABA $(\frac{-\phi}{\phi+3})$ ABB $(\frac{3(\phi+2)}{2(\phi+3)})$			

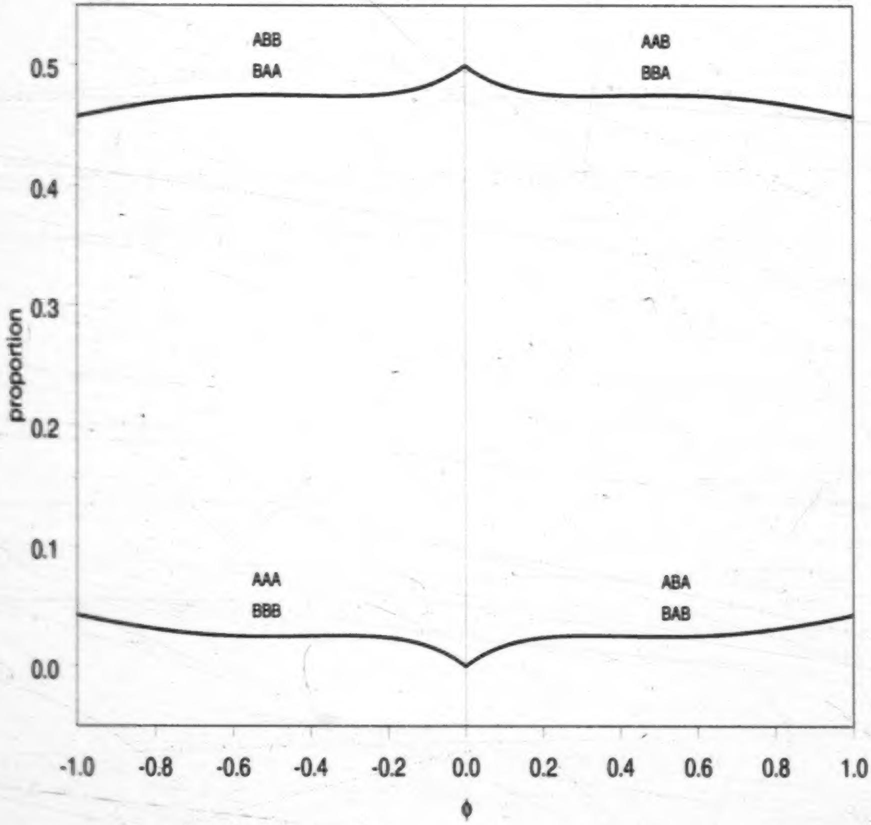

Figure 2.1: Proportions of subjects in each given sequence that makes up optimal three-period designs for estimating treatment effects, when $\rho = 0.0$.

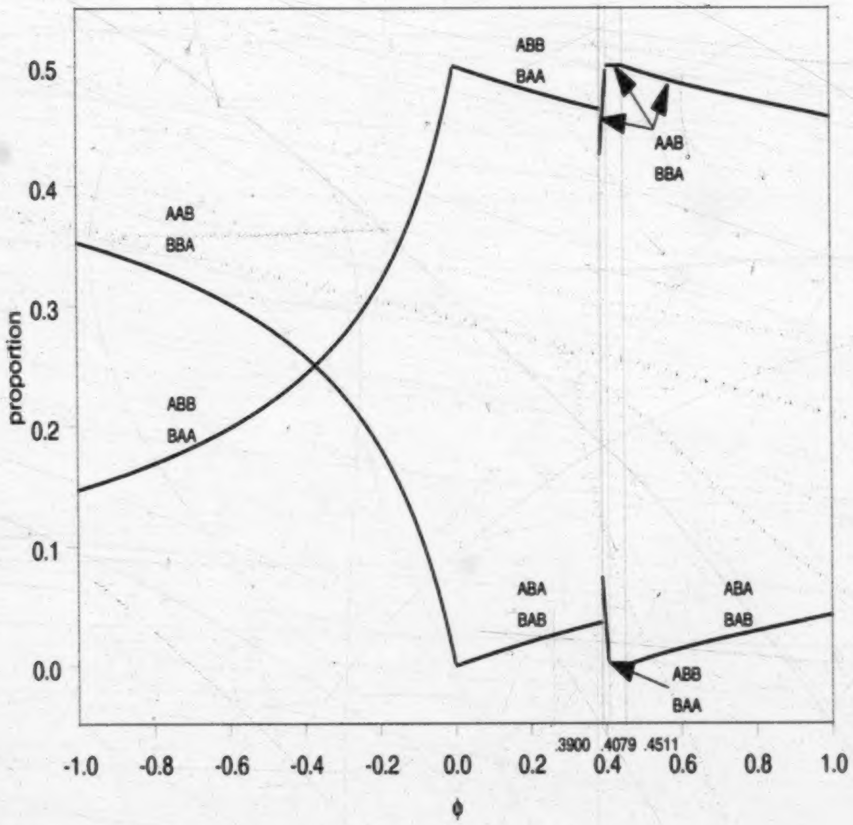

Figure 2.2: Proportions of subjects in each given sequence that makes up optimal three-period designs for estimating treatment effects, when $\rho = 0.3$.

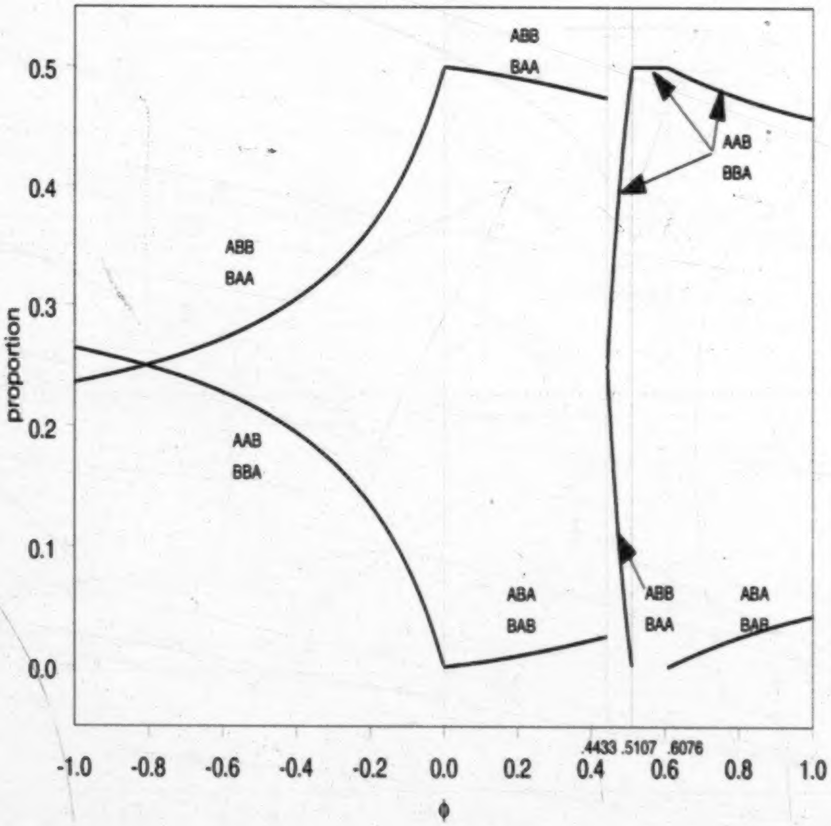

Figure 2.3: Proportions of subjects in each given sequence that makes up optimal three-period designs for estimating treatment effects, when $\rho = 0.7$.

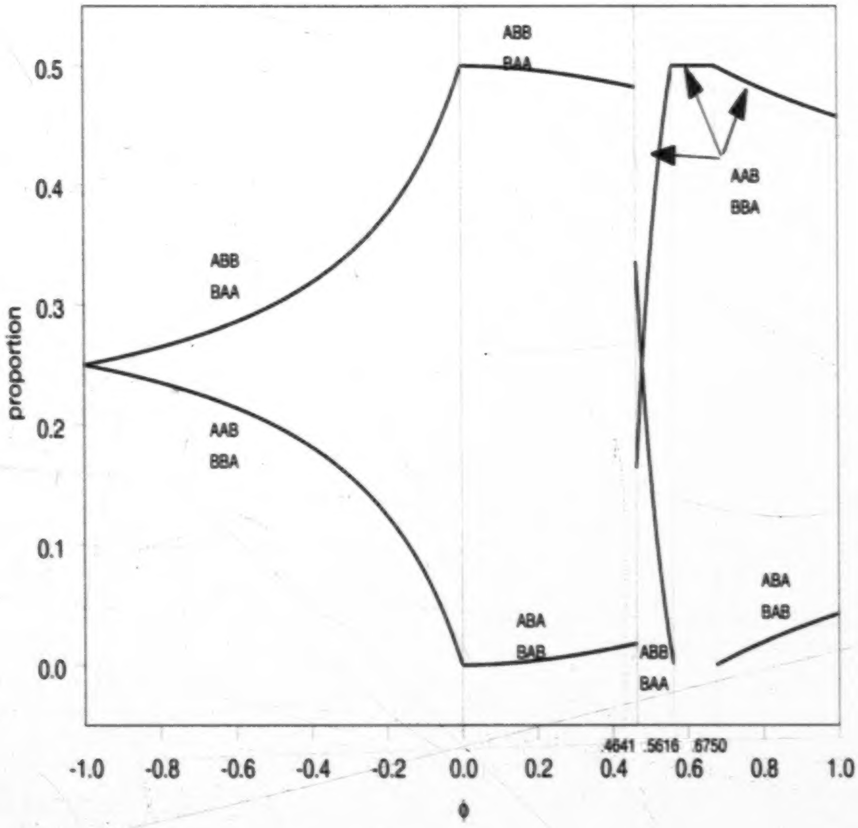

Figure 2.4: Proportions of subjects in each given sequence that makes up optimal three-period designs for estimating treatment effects, when $\rho \to 1$.

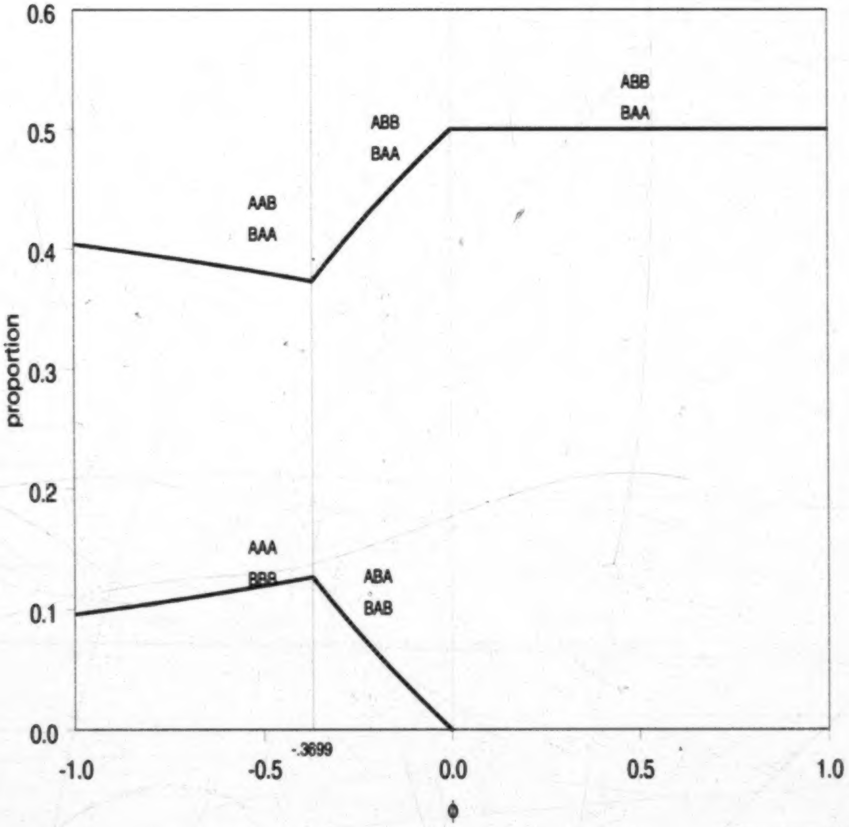

Figure 2.5: Proportions of subjects in each given sequence that makes up optimal three-period designs for estimating residual effects, when $\rho = 0.3$.

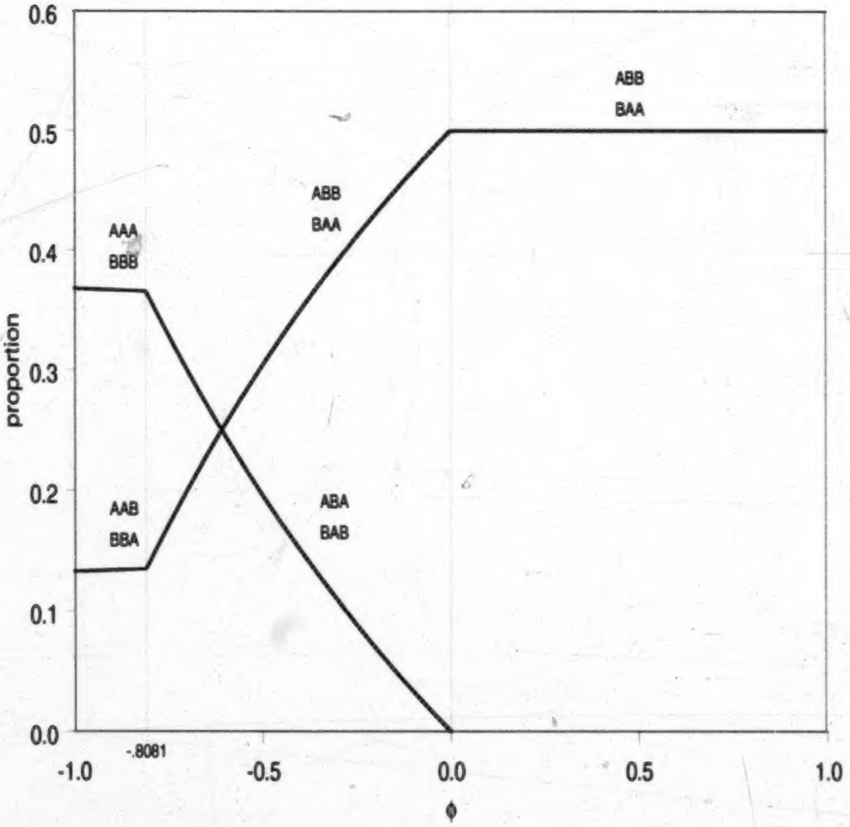

Figure 2.6: Proportions of subjects in each given sequence that makes up optimal three-period designs for estimating residual effects, when $\rho = 0.7$.

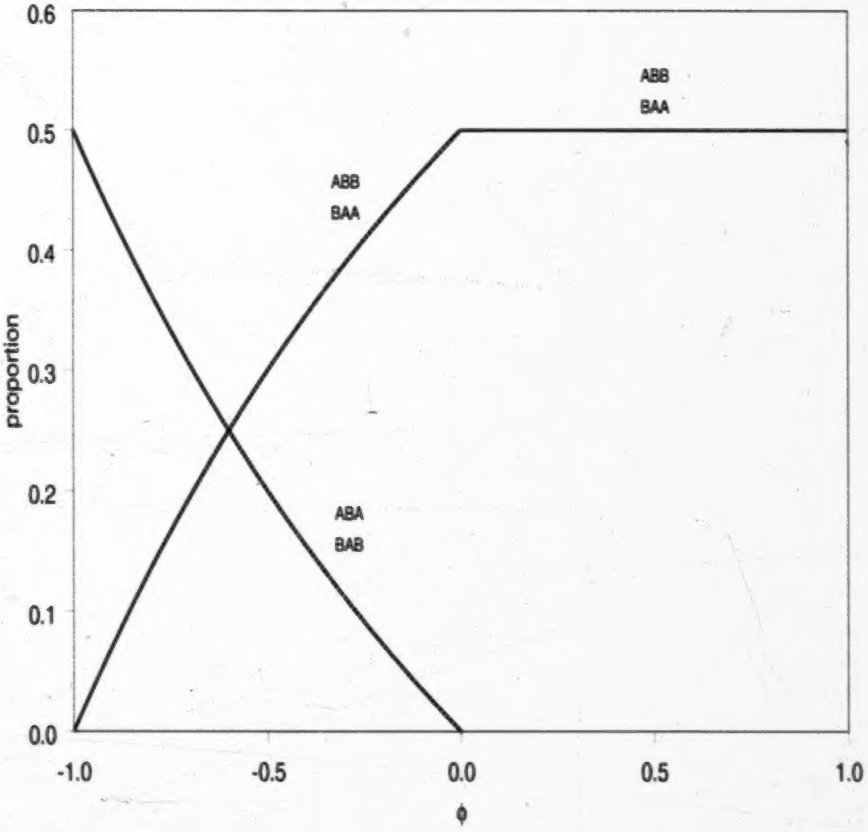

Figure 2.7: Proportions of subjects in each given sequence that makes up optimal three-period designs for estimating residual effects, when $\rho \to 1$.

Table 2.4: Values of ϕ_1 and ϕ_2 at given ρ for four-period optimal designs for estimating treatment effects τ

ρ	ϕ_1	ϕ_2
0	-.4238	.4238
0.1	-.5514	.4211
0.2	-.7431	.4186
0.3		.4165
0.4		.4146
0.5		.4130
0.6		.4115
0.7		.4102
0.8		.4090
0.9		.4079
$\rightarrow 1.0$		.4069

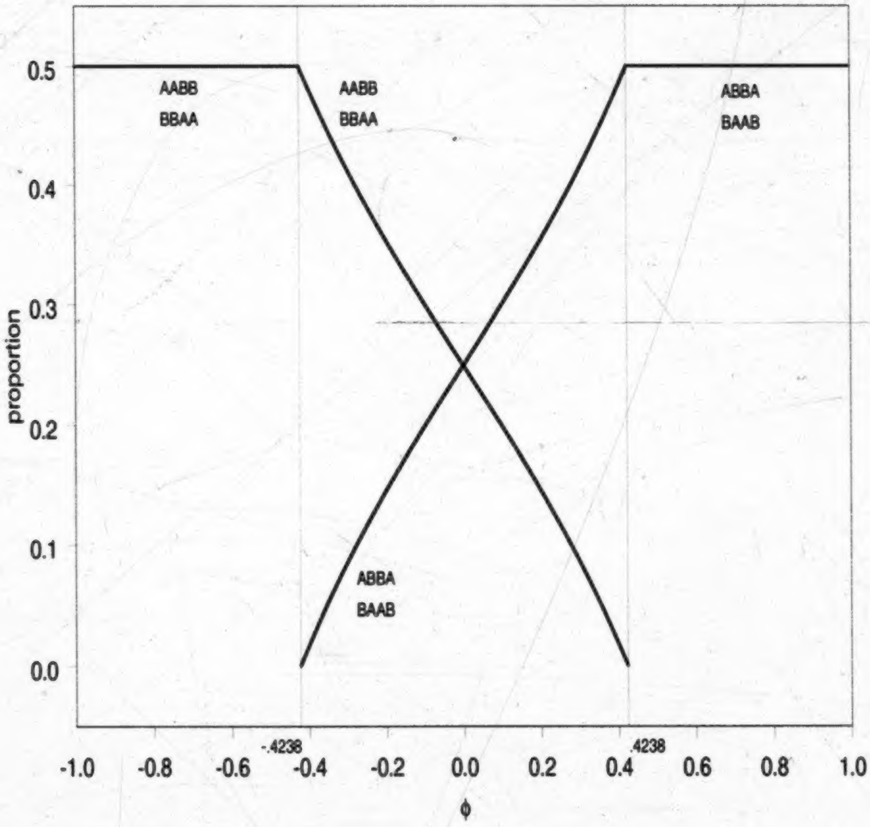

Figure 2.8: Proportions of subjects in each given sequence that makes up optimal four-period designs for estimating treatment effects, when $\rho = 0.0$.

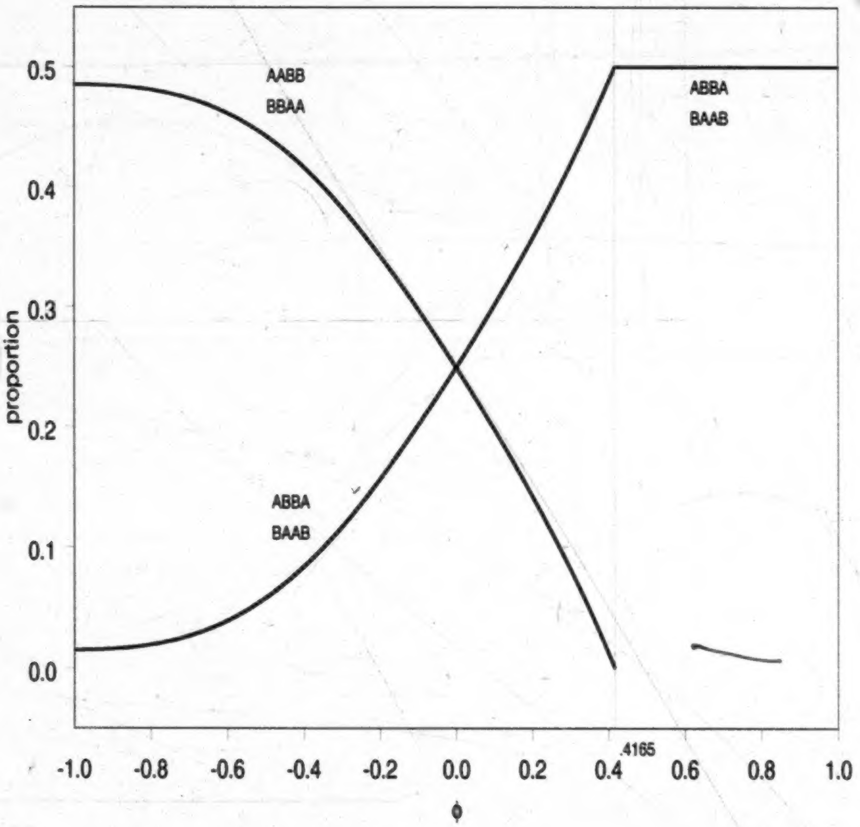

Figure 2.9: Proportions of subjects in each given sequence that makes up optimal four-period designs for estimating treatment effects, when $\rho = 0.3$.

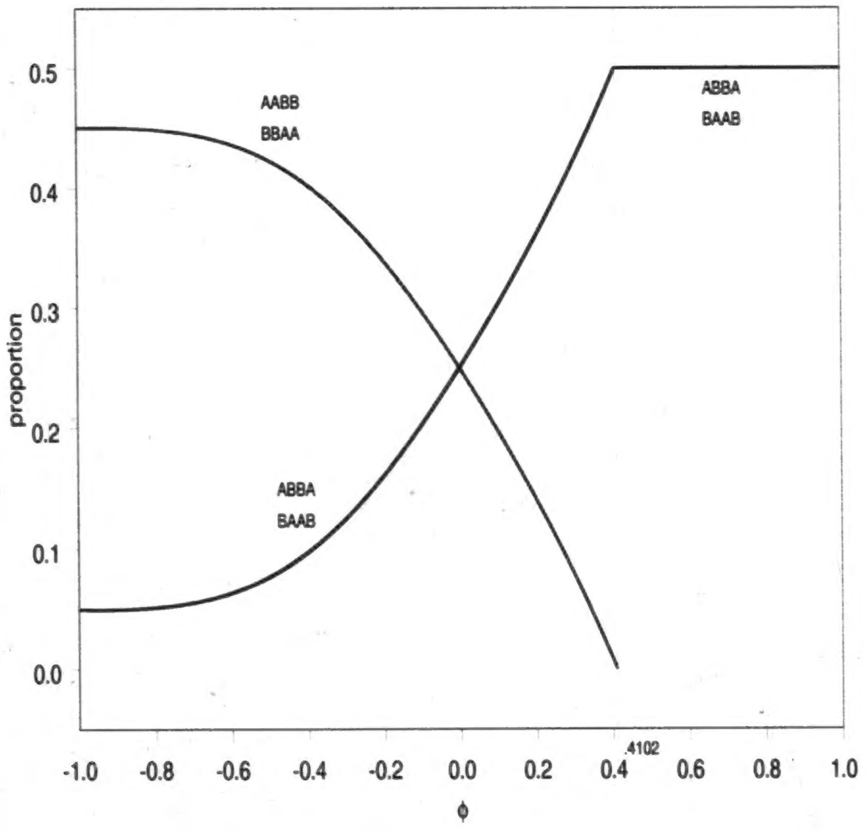

Figure 2.10: Proportions of subjects in each given sequence that makes up optimal four-period designs for estimating treatment effects, when $\rho = 0.7$.

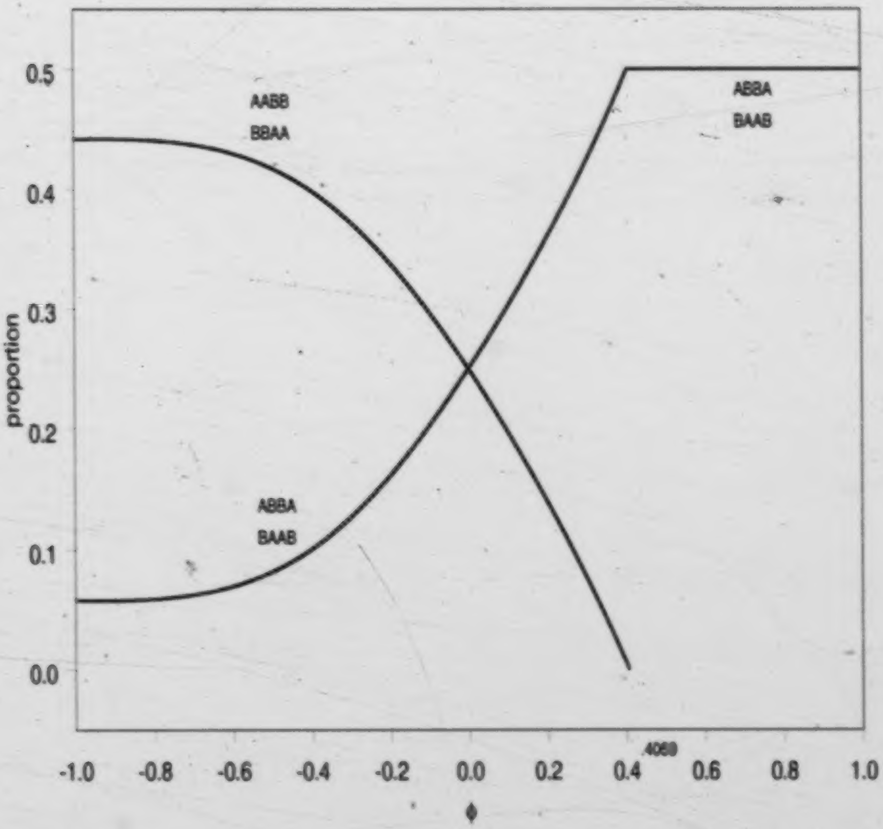

Figure 2.11: Proportions of subjects in each given sequence that makes up optimal four-period designs for estimating treatment effects, when $\rho \to 1$.

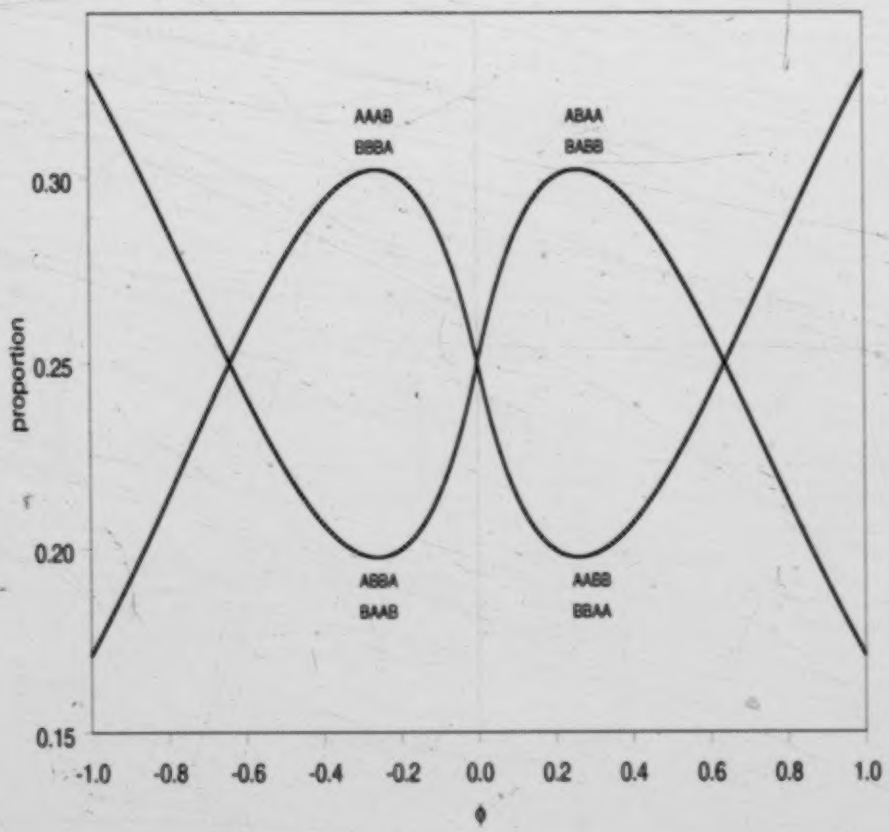

Figure 2.12: Proportions of subjects in each given sequence that makes up optimal four-period designs for estimating residual effects, when $\rho = 0.0$.

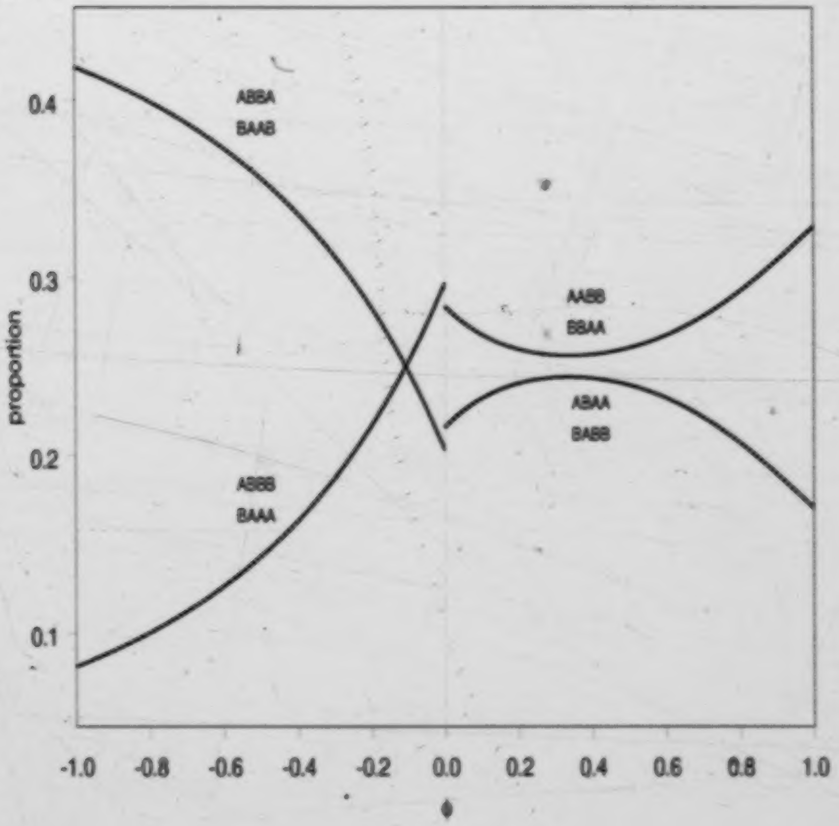

Figure 2.13: Proportions of subjects in each given sequence that makes up optimal four-period designs for estimating residual effects, when $\rho = 0.3$.

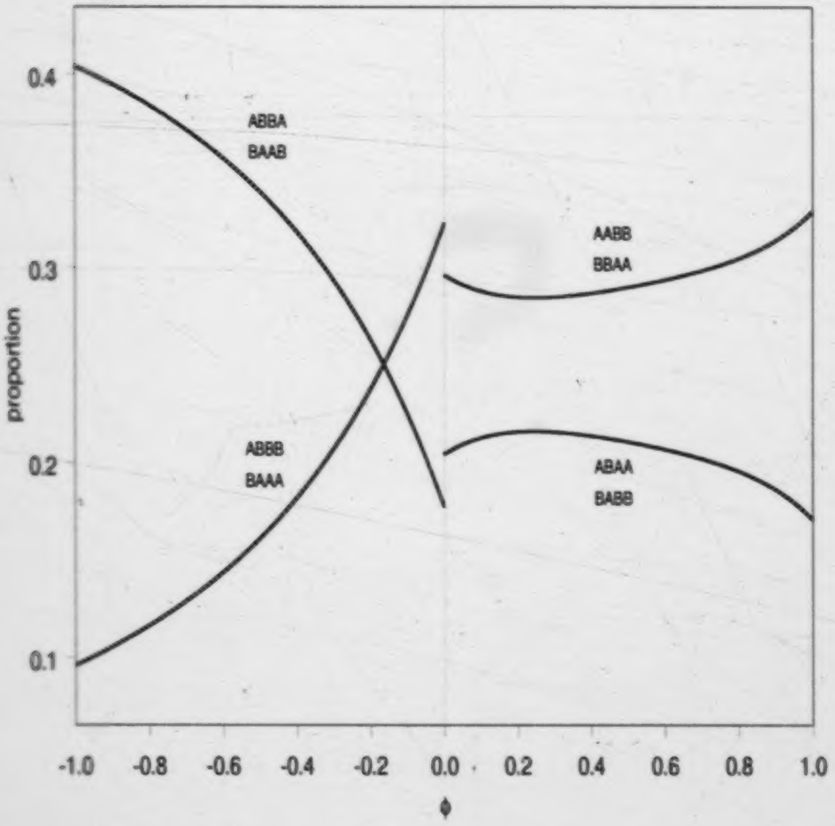

Figure 2.14: Proportions of subjects in each given sequence that makes up optimal four-period designs for estimating residual effects, when $\rho = 0.7$.

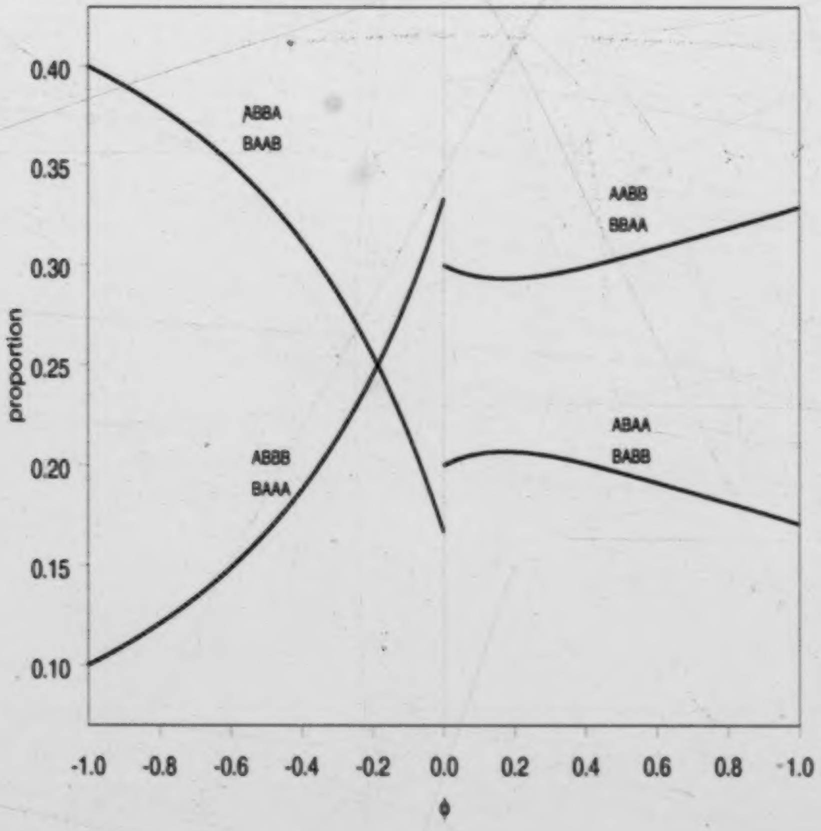

Figure 2.15: Proportions of subjects in each given sequence that makes up optimal four-period designs for estimating residual effects, when $\rho \to 1$.

Table 2.5: Index for two-period, three-period and four-period designs

p	Design	Sequences
2	I	(AB, BA)
	II	(AB, BA), (AA, BB)
	III	(AA, BB)
3	IV	(ABB, BAA)
	V	(ABB, BAA), (AAA, BBB)
	VI	(ABB, BAA), (AAB, BBA)
	VII	(ABB, BAA), (ABA, BAB)
	VIII	(AAA, BBB)
4	IX	(AABB, BBAA)
	X	(ABBA, BAAB)
	XI	(ABBB, BAAA)
	XII	(ABAA, BABB)
	XIII	(ABAB, BABA)
	XIV	(AABB, BBAA), (ABBA, BAAB)
	XV	(AABB, BBAA), (ABAA, BABB)
	XVI	(ABBA, BAAB), (ABBB, BAAA)

Table 2.6: Efficiency of two-period designs relative to the optimal design

Parameter	Design	ϕ								
		-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
$\rho = 0.0$										
τ	*I*	0.26	0.39	0.46	0.49	0.50	0.49	0.46	0.39	0.26
	III	0.26	0.39	0.46	0.49	0.50	0.49	0.46	0.39	0.26
γ	*I*	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10
	II	0.68	0.82	0.92	0.98	1.00	0.98	0.92	0.82	0.68
	III	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
$\rho = 0.3$										
τ	*I*	0.41	0.48	0.50	0.49	0.48	0.45	0.41	0.35	0.24
	III	0.41	0.48	0.50	0.49	0.48	0.45	0.41	0.35	0.24
γ	*I*	0.78	0.63	0.51	0.43	0.35	0.28	0.22	0.16	0.09
	II	0.84	0.97	1.00	0.99	0.96	0.91	0.84	0.76	0.66
	III	0.22	0.37	0.49	0.57	0.65	0.72	0.78	0.84	0.91
$\rho = 0.7$										
τ	*I*	0.50	0.47	0.42	0.38	0.34	0.30	0.27	0.23	0.17
	III	0.50	0.47	0.42	0.38	0.34	0.30	0.27	0.23	0.17
γ	*I*	0.49	0.32	0.24	0.19	0.15	0.12	0.10	0.08	0.05
	II	1.00	0.94	0.86	0.80	0.76	0.72	0.68	0.65	0.60
	III	0.51	0.68	0.76	0.81	0.85	0.88	0.90	0.92	0.95
$\rho \to 1.0$										
τ	*I*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	III	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
γ	*I*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	II	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	III	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 2.7: Efficiency of three-period designs relative to the optimal design when $\rho = 0.0$

Design	ϕ								
	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
τ									
IV	0.99	0.99	1.00	1.00	1.00	0.99	0.98	0.93	0.86
V	0.73	0.75	0.79	0.82	0.83	0.81	0.75	0.64	0.52
VI	0.93	0.96	0.99	1.00	1.00	1.00	0.99	0.96	0.93
VII	0.56	0.64	0.72	0.79	0.83	0.85	0.85	0.82	0.75
VIII	0.23	0.27	0.30	0.32	0.33	0.32	0.28	0.21	0.10
γ									
IV	0.28	0.35	0.47	0.67	1.00	1.00	1.00	1.00	1.00
V	0.37	0.40	0.47	0.61	0.83	0.79	0.75	0.71	0.67
VI	0.67	0.69	0.74	0.84	1.00	0.84	0.74	0.69	0.67
VII	0.28	0.35	0.46	0.61	0.83	0.76	0.70	0.65	0.61
VIII	0.14	0.17	0.21	0.26	0.33	0.3	0.27	0.25	0.25

Table 2.8: Efficiency of three-period designs relative to the optimal design when $\rho = 0.3$

Design	ϕ								
	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
τ									
IV	0.78	0.89	0.95	0.99	1.00	1.00	0.99	0.94	0.86
V	0.51	0.58	0.64	0.68	0.69	0.69	0.67	0.60	0.51
VI	0.99	1.00	1.00	1.00	0.99	0.99	1.00	0.97	0.93
VII	0.57	0.68	0.75	0.80	0.82	0.84	0.85	0.81	0.75
VIII	0.25	0.25	0.26	0.26	0.25	0.24	0.21	0.16	0.09
γ									
IV	0.43	0.65	0.92	0.99	1.00	1.00	1.00	1.00	1.00
V	0.47	0.61	0.78	0.79	0.77	0.74	0.72	0.69	0.66
VI	0.77	0.87	0.98	0.90	0.81	0.74	0.69	0.67	0.67
VII	0.54	0.73	0.93	0.90	0.82	0.75	0.69	0.65	0.61
VIII	0.36	0.39	0.43	0.38	0.33	0.30	0.27	0.26	0.25

Table 2.9: Efficiency of three-period designs relative to the optimal design when $\rho = 0.7$

Design	ϕ								
	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
τ									
IV	0.61	0.83	0.94	0.99	1.00	1.00	1.00	0.95	0.86
V	0.41	0.48	0.53	0.55	0.57	0.57	0.57	0.54	0.48
VI	1.00	1.00	0.99	0.99	0.98	0.98	0.99	0.97	0.93
VII	0.56	0.69	0.77	0.80	0.82	0.82	0.83	0.80	0.75
VIII	0.13	0.13	0.12	0.12	0.11	0.10	0.10	0.08	0.06
γ									
IV	0.61	0.83	0.94	0.99	1.00	1.00	1.00	1.00	1.00
V	0.69	0.72	0.73	0.72	0.70	0.68	0.67	0.66	0.65
VI	0.99	0.88	0.80	0.74	0.69	0.66	0.65	0.65	0.66
VII	0.94	1.00	0.96	0.89	0.82	0.74	0.69	0.64	0.61
VIII	0.70	0.55	0.45	0.38	0.33	0.30	0.27	0.26	0.25

Table 2.10: Efficiency of three-period designs relative to the optimal design when ρ approaches 1.0

Design	ϕ								
	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
τ									
IV	0.54	0.80	0.93	0.98	1.00	1.00	1.00	0.95	0.86
V	0.37	0.44	0.48	0.49	0.50	0.50	0.50	0.48	0.44
VI	1.00	0.99	0.99	0.98	0.97	0.97	0.99	0.98	0.93
VII	0.56	0.70	0.77	0.80	0.81	0.82	0.82	0.80	0.74
VIII	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
γ									
IV	0.54	0.80	0.93	0.98	1.00	1.00	1.00	1.00	1.00
V	0.63	0.68	0.69	0.68	0.67	0.65	0.64	0.63	0.63
VI	0.86	0.77	0.72	0.68	0.65	0.63	0.63	0.64	0.66
VII	0.93	1.00	0.96	0.89	0.81	0.74	0.69	0.64	0.61
VIII	0.71	0.56	0.45	0.38	0.33	0.30	0.27	0.26	0.25

Table 2.11: Efficiency of four-period designs relative to the optimal design when $\rho = 0.0$

	ϕ								
Design	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
τ									
IX	1.00	1.00	1.00	0.97	0.92	0.84	0.75	0.66	0.60
X	0.60	0.66	0.75	0.84	0.92	0.97	1.00	1.00	1.00
XI	0.68	0.73	0.81	0.88	0.92	0.88	0.77	0.64	0.53
XII	0.60	0.67	0.77	0.87	0.92	0.91	0.85	0.74	0.65
XIII	0.07	0.14	0.20	0.24	0.25	0.24	0.21	0.18	0.15
XIV	0.84	0.88	0.94	0.99	1.00	0.99	0.94	0.88	0.84
XV	0.81	0.85	0.92	0.98	1.00	0.98	0.90	0.81	0.72
XVII	0.76	0.83	0.91	0.98	1.00	0.98	0.91	0.83	0.76
γ									
IX	0.77	0.75	0.75	0.80	0.92	0.84	0.82	0.83	0.86
X	0.86	0.83	0.82	0.84	0.92	0.80	0.75	0.75	0.77
XI	0.53	0.54	0.61	0.73	0.92	0.88	0.84	0.80	0.76
XII	0.22	0.30	0.44	0.64	0.92	0.94	0.94	0.91	0.87
XIII	0.19	0.19	0.21	0.23	0.25	0.21	0.18	0.16	0.14
XIV	0.84	0.83	0.84	0.89	1.00	0.89	0.84	0.83	0.84
XV	0.51	0.54	0.62	0.77	1.00	1.00	1.00	1.00	1.00
XVII	0.77	0.78	0.81	0.88	1.00	0.88	0.81	0.78	0.77

Table 2.12: Efficiency of four-period designs relative to the optimal design when $\rho = 0.3$

Design	ϕ								
	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
					τ				
IX	1.00	1.00	0.99	0.96	0.91	0.84	0.76	0.67	0.60
X	0.52	0.62	0.72	0.82	0.91	0.97	1.00	1.00	1.00
XI	0.36	0.46	0.59	0.71	0.80	0.81	0.75	0.63	0.53
XII	0.44	0.49	0.55	0.63	0.69	0.73	0.73	0.69	0.63
XIII	0.07	0.13	0.17	0.20	0.21	0.21	0.19	0.17	0.15
XIV	0.86	0.91	0.95	0.99	1.00	0.99	0.94	0.88	0.84
XV	0.80	0.82	0.86	0.90	0.92	0.91	0.86	0.79	0.72
XVI	0.54	0.65	0.76	0.85	0.92	0.93	0.89	0.82	0.76
					γ				
IX	0.51	0.61	0.72	0.83	0.91	0.86	0.84	0.85	0.87
X	0.97	0.96	0.94	0.93	0.91	0.80	0.76	0.75	0.77
XI	0.57	0.66	0.76	0.87	0.95	0.89	0.84	0.80	0.76
XII	0.19	0.29	0.45	0.64	0.82	0.85	0.87	0.87	0.86
XIII	0.16	0.18	0.19	0.20	0.21	0.18	0.16	0.15	0.13
XIV	0.80	0.85	0.91	0.97	1.00	0.90	0.85	0.84	0.84
XV	0.39	0.50	0.65	0.83	1.00	1.00	1.00	1.00	1.00
XVI	0.94	0.96	0.98	1.00	1.00	0.87	0.81	0.78	0.77

Table 2.13: Efficiency of four-period designs relative to the optimal design when $\rho = 0.7$

	ϕ								
Design	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
					τ				
IX	0.99	0.99	0.98	0.96	0.91	0.84	0.76	0.67	0.61
X	0.51	0.60	0.71	0.82	0.91	0.97	1.00	1.00	1.00
XI	0.29	0.39	0.52	0.65	0.75	0.77	0.72	0.63	0.53
XII	0.39	0.43	0.48	0.53	0.58	0.62	0.63	0.62	0.60
XIII	0.07	0.13	0.16	0.18	0.19	0.19	0.17	0.16	0.14
XIV	0.88	0.91	0.95	0.99	1.00	0.99	0.94	0.88	0.84
XV	0.80	0.81	0.84	0.87	0.88	0.87	0.82	0.76	0.71
XVI	0.50	0.60	0.71	0.81	0.88	0.90	0.87	0.82	0.76
					γ				
IX	0.43	0.55	0.67	0.80	0.91	0.87	0.86	0.86	0.87
X	0.96	0.95	0.93	0.92	0.91	0.80	0.76	0.76	0.78
XI	0.54	0.65	0.77	0.88	0.96	0.89	0.83	0.79	0.76
XII	0.17	0.27	0.41	0.59	0.75	0.78	0.80	0.82	0.83
XIII	0.15	0.16	0.18	0.19	0.19	0.16	0.15	0.14	0.13
XIV	0.76	0.82	0.89	0.96	1.00	0.91	0.86	0.84	0.85
XV	0.35	0.46	0.62	0.81	0.99	1.00	1.00	1.00	0.99
XVI	0.94	0.97	0.99	1.00	0.99	0.87	0.80	0.78	0.77

Table 2.14: Efficiency of four-period designs relative to the optimal design when ρ approaches 1.0

Design	ϕ								
	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
τ									
IX	0.98	0.98	0.98	0.95	0.91	0.84	0.76	0.68	0.61
X	0.50	0.60	0.71	0.81	0.91	0.97	1.00	1.00	1.00
XI	0.27	0.37	0.50	0.63	0.73	0.76	0.71	0.62	0.53
XII	0.37	0.41	0.46	0.50	0.55	0.57	0.58	0.57	0.56
XIII	0.07	0.13	0.16	0.18	0.18	0.18	0.16	0.15	0.14
XIV	0.88	0.91	0.95	0.99	1.00	0.98	0.94	0.88	0.84
XV	0.80	0.81	0.83	0.86	0.87	0.85	0.80	0.75	0.70
XVI	0.49	0.59	0.70	0.80	0.87	0.89	0.87	0.81	0.77
γ									
IX	0.41	0.53	0.66	0.79	0.91	0.88	0.86	0.87	0.88
X	0.95	0.94	0.93	0.92	0.91	0.80	0.76	0.76	0.78
XI	0.53	0.65	0.78	0.89	0.97	0.89	0.83	0.79	0.75
XII	0.16	0.26	0.40	0.57	0.73	0.74	0.76	0.77	0.79
XIII	0.15	0.16	0.17	0.18	0.18	0.16	0.14	0.13	0.12
XIV	0.75	0.82	0.89	0.95	1.00	0.91	0.86	0.85	0.85
XV	0.34	0.45	0.62	0.81	0.99	0.99	0.99	0.99	0.99
XVI	0.95	0.97	0.99	1.00	0.99	0.86	0.80	0.77	0.77

Chapter 3

Response Adaptive Repeated Measurement Designs

3.1 Introduction

As discussed in the previous chapter, much of the work on constructing optimal designs is based on two assumptions: that subjects are associated with a common covariance matrix and that the covariance structure is known. Consequently, these optimal designs are rather sensitive to violations of these assumptions. These designs may become inefficient when the data does not support the assumed covariance matrix. For instance, the design using the sequences ABB and BAA is optimal for estimating the contrast of residual effects

γ when the covariance matrix is first-order autoregressive with the parameter $\phi \geq 0$ and no subject effects (see Chapter 2). However, if the data indicate that the value of ϕ is $-.8$, then the efficiency of that design is only .28 relative to the optimal design for that situation.

In Chapter 2, we recommended some designs that are robust against the violation of model assumptions. We have to point out, however, that this robustness is restricted to the variation of some parameters determining the covariance matrix, assuming that the form of a covariance matrix is known. These designs may not be robust in other situations. Methods are needed to construct designs without having to assume an error structure because, in practice, the error structure is not known before an experiment but after a number of subjects are entered into the experiment.

For Design I (defined in Chapter 2, Table 2.5), Cook (1995, 1996) provided guidelines for terminating the accession of subjects in terms of power and type I error. The attractions of Cook's method are: it utilizes the estimates of the assumed common covariance matrix obtained adaptively to the subject response, it allows the use of subjects with only the first period observed and it results in savings in terms of the smaller expected measurements required as compared with the fixed-sample-size cross-over design. Not restricted to one specific design, Kushner (2000) proposed a method for constructing general cross-over designs adaptively, that is, relaxing the assumption of a known error structure but still assuming homogeneity among the study subjects. Rather than terminating the adaptive procedure to minimize the expected sample size, as is the case in the traditional sequential trial design (Armitage 1975), the focus was on improving the precision of the estimators of parameters, similarly as in Chapter 2.

However, Kushner's approach leaves something to be desired in that the covariance matrices for the within-subject measurement error may not be the same for all subjects. For example, the subjects can exhibit a wide range of variation in responses, depending on their demographic or prognostic factors. Here, we aim to relax further the usual assumption about the covariance matrix by removing the homogeneity restriction.

In this chapter, we propose a method of sequentially allocating a small number of new subjects to treatment sequences, when the covariance matrices for the subjects are unknown and heterogeneous. Practically, we can handle only a finite number of different covariance matrices in an experiment. Further, we assume some information is available on who is associated with each of the covariance matrices being considered.

As a simple example, consider a two-treatment two-period repeated measurement design with the sequences AB and BA. Patients are assigned to a group that receives a new drug (A) followed by a placebo (B), or to a group that receives B first and then A. Older patients may be associated with a different covariance matrix than younger patients. Also, a patient's chances of being associated with a particular covariance matrix will depend on the progression of a disease. Usually, this type of information can be known only approximately. It can be obtained from previous studies or estimated from a pilot project. We derive the allocation rules by minimizing a loss function of the increment of information for each new subject or a small group of new subjects. Some adjustments will also be considered for situations when only probabilistic information is available about a particular patient's membership to an error structure. The allocation rules are applied sequentially to new subjects, and the parameters involved are updated, using the subjects already in the experiment, and aiming eventually for efficient estimators of the treatment effects.

The idea is similar to the approach of traditional adaptive designs in clinical trials—the accrued information on responses to treatment is used to assign new subjects (Bather, 1985; Armitage, 1985; Rosenberger and Lachin, 1993). However, our interest is in building adaptive repeated measurement designs, as opposed to the typical randomized designs. Further, our allocation rules are derived to assign subjects to treatment sequences that contribute most to the information matrix for the parameters of interest, rather than to a better performed treatment, as in traditional adaptive designs (Wei and Durham, 1978; Wei, 1979; Rosenberger, 1995). Finally, the proposed adaptive design is intended to obtain efficient estimators for parameters of interest, rather than to minimize the proportion of subjects assigned to an inferior treatment (Zelen, 1969; Simon, 1977, 1991; Pocock, 1979; Bartlett et al., 1985). The important contribution of this chapter is our efforts to incorporate heterogeneity among study subjects into the construction of adaptive designs.

This chapter is organized as follows. In Section 2, we introduce the model and present the information matrix for the parameters of interest. In Section 3, we develop the allocation rules for constructing the designs. In Section 4, we demonstrate the efficiency of the resulting designs as compared to the optimal designs, in selected situations. In Section 5, we summarize our approach and provide concluding remarks.

3.2 Information matrix for treatment effects

Consider model (2.2.1) with unknown and heterogeneous covariance matrices associated with the study subjects, where $\mathbf{V}_{jk}$ are not all equal, for $k = 1, \ldots, s$, $j = 1, \ldots, N_k$. Here, we assume that $\mathbf{V}_{jk}$ can be grouped into a finite set, consisting of $\mathbf{V}_1, \ldots, \mathbf{V}_M$. The subject effects in the model can be fixed or random.

For model (2.2.1) with random subject effects, we have

$$\mathbf{V}_{jk}^{-1} = \mathbf{\Sigma}_{jk}^{-1} - \sigma_\xi^2 \mathbf{\Sigma}_{jk}^{-1} \mathbf{1}_{[p]} \mathbf{1}_{[p]}^T \mathbf{\Sigma}_{jk}^{-1} / (1 + \sigma_\xi^2 \mathbf{1}_{[p]}^T \mathbf{\Sigma}_{jk}^{-1} \mathbf{1}_{[p]}),$$

while for model (2.2.1) with fixed subject effects, it becomes

$$\mathbf{V}_{jk}^{-1} = \mathbf{\Sigma}_{jk}^{-1} - \mathbf{\Sigma}_{jk}^{-1} \mathbf{1}_{[p]} \mathbf{1}_{[p]}^T \mathbf{\Sigma}_{jk}^{-1} / \mathbf{1}_{[p]}^T \mathbf{\Sigma}_{jk}^{-1} \mathbf{1}_{[p]}.$$

Define $\overline{\mathbf{X}}^\tau$ and $\overline{\mathbf{X}}^\gamma$ to be $\frac{1}{N}(\overline{\mathbf{V}^{-1}})^{-1} \sum_{jk} \mathbf{V}_{jk}^{-1} \mathbf{X}_k^\tau$ and $\frac{1}{N}(\overline{\mathbf{V}^{-1}})^{-1} \sum_{jk} \mathbf{V}_{jk}^{-1} \mathbf{X}_k^\gamma$, respectively, where $\overline{\mathbf{V}^{-1}} = \sum_{jk} N_k \mathbf{V}_{jk}^{-1}/N = \sum_{km} n_{km} \mathbf{V}_m^{-1}/N$ with n_{km} being the number of subjects in sequence k associated with covariance matrix $\mathbf{V}_m$, $m = 1, \ldots, M$, $k = 1, \ldots, s$. This leads us to the following Lemmas.

Lemma 3.2.1 *The information matrix for* $(\boldsymbol{\tau}^T, \boldsymbol{\gamma}^T)^T$ *is*

$$\mathbf{I}(\boldsymbol{\tau}, \boldsymbol{\gamma}) = \begin{bmatrix} \mathbf{I}_{11} & \mathbf{I}_{12} \\ \mathbf{I}_{21} & \mathbf{I}_{22} \end{bmatrix} \tag{3.2.1}$$

where $\mathbf{I}_{11} = \sum_{k=1}^{s} \sum_{j=1}^{N_k} (\mathbf{X}_k^\tau - \overline{\mathbf{X}}^\tau)^T \mathbf{V}_{jk}^{-1} (\mathbf{X}_k^\tau - \overline{\mathbf{X}}^\tau)$, $\mathbf{I}_{12} = \sum_{k=1}^{s} \sum_{j=1}^{N_k} (\mathbf{X}_k^\tau - \overline{\mathbf{X}}^\tau)^T \mathbf{V}_{jk}^{-1} (\mathbf{X}_k^\gamma - \overline{\mathbf{X}}^\gamma)$, $\mathbf{I}_{21} = \mathbf{I}_{12}^T$ *and* $\mathbf{I}_{22} = \sum_{k=1}^{s} \sum_{j=1}^{N_k} (\mathbf{X}_k^\gamma - \overline{\mathbf{X}}^\gamma)^T \mathbf{V}_{jk}^{-1} (\mathbf{X}_k^\gamma - \overline{\mathbf{X}}^\gamma)$.

PROOF. When the subject effects are random, the information matrix for

$\boldsymbol{\beta} = (\mu, \boldsymbol{\pi}^T, \boldsymbol{\tau}^T, \boldsymbol{\gamma}^T)^T$ is

$$\mathbf{I}(\boldsymbol{\beta}) = \mathbf{X}^T Diag\{\mathbf{V}_{jk}^{-1}\}\mathbf{X}$$

where $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, $\mathbf{X}_1 = (\mathbf{1}_N \otimes \mathbf{1}_{[p]}, \mathbf{1}_{[\mathbf{N}]} \otimes \mathbf{I}_{[\mathbf{p}]})$ and $\mathbf{X}_2 = (\mathbf{1}_{N_1}^T \otimes \mathbf{X}_{21}^T, \ldots, \mathbf{1}_{N_s}^T \otimes \mathbf{X}_{2s}^T)^T$, with $\mathbf{X}_{2k}$ defined as in Chapter 2. Then, the information for $(\boldsymbol{\tau}^T, \boldsymbol{\gamma}^T)^T$ is

$$\begin{aligned} \mathbf{I}(\boldsymbol{\tau}, \boldsymbol{\gamma}) &= \mathbf{X}_2^T Diag\{\mathbf{V}_{jk}^{-1}\}\mathbf{X}_2 \\ &\quad - \mathbf{X}_2^T Diag\{\mathbf{V}_{jk}^{-1}\}\mathbf{X}_1(\mathbf{X}_1^T Diag\{\mathbf{V}_{jk}^{-1}\}\mathbf{X}_1)^{-}\mathbf{X}_1^T Diag\{\mathbf{V}_{jk}^{-1}\}\mathbf{X}_2. \end{aligned}$$

Note that

$$\mathbf{X}_1(\mathbf{X}_1^T Diag\{\mathbf{V}_{jk}^{-1}\}\mathbf{X}_1)^{-}\mathbf{X}_1^T = \frac{1}{N}\mathbf{1}_N \mathbf{1}_N^T \otimes \overline{\mathbf{V}^{-1}}^{-1}.$$

Therefore,

$$\mathbf{I}(\boldsymbol{\tau}, \boldsymbol{\gamma}) = \sum_{jk} \mathbf{X}_{2k}^T \mathbf{V}_{jk}^{-1} \mathbf{X}_{2k} - \frac{1}{N} \sum_{jk} \mathbf{X}_{2k}^T \mathbf{V}_{jk}^{-1} (\overline{\mathbf{V}^{-1}})^{-1} \sum_{jk} \mathbf{V}_{jk}^{-1} \mathbf{X}_{2k},$$

and it follows that

$$\begin{aligned} \mathbf{I}_{11} &= \sum_{jk} \mathbf{X}_k^{\tau T} \mathbf{V}_{jk}^{-1} \mathbf{X}_k^{\tau} - \frac{1}{N} \sum_{jk} \mathbf{X}_k^{\tau T} \mathbf{V}_{jk}^{-1} (\overline{\mathbf{V}^{-1}})^{-1} \sum_{jk} \mathbf{V}_{jk}^{-1} \mathbf{X}_k^{\tau} \\ &= \sum_{jk} \mathbf{X}_k^{\tau T} \mathbf{V}_{jk}^{-1} \mathbf{X}_k^{\tau} - N \overline{\mathbf{X}}^{\tau T} \overline{\mathbf{V}^{-1} \mathbf{X}^{\tau}} \end{aligned}$$

$$= \sum_{jk} (\mathbf{X}_k^\tau - \overline{\mathbf{X}}^\tau)^T \mathbf{V}_{jk}^{-1} (\mathbf{X}_k^\tau - \overline{\mathbf{X}}^\tau),$$

using $\overline{\mathbf{X}}^\tau$ and $\overline{\mathbf{V}^{-1}}$ as defined earlier. Similarly, $\mathbf{I}_{12}$ and $\mathbf{I}_{22}$ can be obtained to give Lemma 3.2.1.

When the subject effects are fixed, the information matrix for $\boldsymbol{\beta}$ becomes

$$\begin{aligned} \mathbf{I}(\boldsymbol{\beta}) &= \mathbf{X}^T Diag\{\boldsymbol{\Sigma}_{jk}^{-1}\}\mathbf{X} \\ &\quad -\mathbf{X}^T Diag\{\boldsymbol{\Sigma}_{jk}^{-1}\}\mathbf{I}_{[N]} \otimes \mathbf{1}_{[p]} Diag\{(\mathbf{1}_{[p]}^T \boldsymbol{\Sigma}_{jk} \mathbf{1}_{[p]})^{-1}\}\mathbf{I}_{[N]} \otimes \mathbf{1}_{[p]}^T Diag\{\boldsymbol{\Sigma}_{jk}^{-1}\}\mathbf{X} \\ &= \mathbf{X}^T Diag\{\mathbf{V}_{jk}^{-1}\}\mathbf{X}, \end{aligned}$$

where $\mathbf{V}_{jk} = \boldsymbol{\Sigma}_{jk}^{-1} - \frac{\boldsymbol{\Sigma}_{jk}^{-1}\mathbf{1}_{[p]}\mathbf{1}_{[p]}^T\boldsymbol{\Sigma}_{jk}^{-1}}{\mathbf{1}_{[p]}^T\boldsymbol{\Sigma}_{jk}^{-1}\mathbf{1}_{[p]}^T}$. Then, the results follow by similar arguments.

□

Lemma 3.2.2 *The row and column sums of* $\mathbf{I}_{11}, \mathbf{I}_{12}, \mathbf{I}_{21}$ *and* $\mathbf{I}_{22}$ *in Lemma 3.2.1 are zero.*

PROOF. Since in each period, one and only one treatment must be assigned, we have $\mathbf{X}_k^\tau \mathbf{1}_{[t]} = \mathbf{1}_{[p]}$, for all k. Noting that $\mathbf{x}_{ik}^\gamma = \mathbf{x}_{i-1,k}^\tau$ for $i = 2, \ldots, p$ and $\mathbf{x}_{1k}^\gamma = \mathbf{0}$, we have that $\mathbf{X}_k^\gamma \mathbf{1}_{[t]} = \mathbf{1}_{[p]}^*$ for all k, where $\mathbf{1}_{[p]}^*$ is the same as $\mathbf{1}_{[p]}$ except for the first element, which is 0. Thus, we have $\overline{\mathbf{X}}_k^\tau \mathbf{1}_{[t]} = \mathbf{1}_{[p]}$ and $\overline{\mathbf{X}}_k^\gamma \mathbf{1}_{[t]} = \mathbf{1}_{[p]}^*$. Therefore, Lemma 3.2.2 follows.

□

Definition 3.2.1 *A repeated measurement design d is a design specified by a vector* $\boldsymbol{n} = \{\boldsymbol{n}_k\}$*, where* $\boldsymbol{n}_k = (n_{k1}, \ldots, n_{kM})$ *is the number of subjects in treatment sequence k associated with each of the M covariance matrices for* $k = 1, \ldots, t^p$*. Then, there exists a 1-1 mapping between d and* $\boldsymbol{n}$ $(d \leftrightarrow \boldsymbol{n})$*.*

Definition 3.2.2 *Design d is symmetric if it remains unchanged by a permutation of* $(1, \ldots, t)$*.*

As defined in Chapter 2, sequences are dual to each other if one sequence can be turned into other sequences by permuting the treatments. Hence, if a symmetric design includes a certain sequence, it also includes the dual sequences of that sequence.

Example of two-treatment symmetric designs: For two-period designs, a design with the sequence AB and its dual is a symmetric design. For three-period designs, a design with the sequences ABB, AAA and their duals is a symmetric two-treatment three-period design.

Example of three-treatment symmetric designs: For two-period designs, a design with the sequences AA, AB and their duals (BB, CC; BA, AC, BC, CA, CB) is a symmetric design. Similarly, for three-period designs, a design with the sequences ABB, AAB and their duals (ACC, BAA, BCC, CAA, CBB; AAC, BBA, BBC, CCA, CCB) is a symmetric design.

Theorem 3.2.1 *If the symmetric design is also dual-balanced, i.e.,* $n_{km} =$

$n_{k^* m}$ for $k = 1, \ldots, s$, k^* is the dual sequence of k, and $m = 1, \ldots, M$, then the information matrix (3.2.1) can be rewritten as

$$\mathbf{I}(\boldsymbol{\tau}, \boldsymbol{\gamma}) = \frac{1}{t(t-1)} \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \otimes (t\mathbf{I}_{[t]} - \mathbf{1}_{[t]}\mathbf{1}_{[t]}^T) \tag{3.2.2}$$

where $a_{ll'} = tr(\mathbf{I}_{ll'}) = \sum_{k=1}^{s} \sum_{m=1}^{M} n_{km} a_{ll'}^{km}$, $\mathbf{I}_{ll'}$ are the ll'-th components of $\mathbf{I}(\boldsymbol{\tau}, \boldsymbol{\gamma})$ in Lemma 3.2.1, $l = 1, 2, l' = 1, 2$, and $a_{11}^{km} = \sum_{i,i'=1}^{p} \delta_k^{ii'} (\mathbf{V}_m^{-1})_{ii'} - v_1/t$, $a_{12}^{km} = \sum_{i=1}^{p} \sum_{i'=1}^{p-1} \delta_k^{ii'} (\mathbf{V}_m^{-1})_{i,i'+1} - v_2/t$, $a_{22}^{km} = \sum_{i=1}^{p-1} \sum_{i'=1}^{p-1} \delta_k^{ii'} (\mathbf{V}_m^{-1})_{i+1,i'+1} - v_3/t$, $v_1 = \mathbf{1}_{[p]}^T \overline{\mathbf{V}^{-1}} \mathbf{1}_{[p]}$, $v_2 = \mathbf{1}_{[p]}^T \overline{\mathbf{V}^{-1}} \mathbf{1}_{[p]}^*$, $v_3 = \mathbf{1}_{[p]}^{*\,T} \overline{\mathbf{V}^{-1}} \mathbf{1}_{[p]}^*$, $\delta_k^{ii'} = 1$ if the i^{th} row and i'^{th} row of $\mathbf{X}_k^{\tau}$ are the same, and 0, otherwise.

Proof. Denote the number of distinct treatments in sequence k as l_k. For any sequence k in a symmetric design d, the design matrices of its $l_k! C_t^{l_k}$ dual sequences (including itself) are generated by permuting the column of $\mathbf{X}_k^{\tau}$ and $\mathbf{X}_k^{\gamma}$. If one treatment in a certain period in sequence k is fixed, there are totally $(l_k - 1)! C_{t-1}^{l_k - 1}$ dual sequences of sequence k including itself.

For instance, in a three-period three-treatment design, the sequences ABC, ACB, BAC, BCA, CAB, CBA are dual sequences, and the number of duals is $3! C_3^3 = 6$. If treatment A is fixed in the first period, sequences ABC and ACB are dual sequences and the number of such dual sequences is $(2-1)! C_2^1 = 2$. Similarly, the sequences BCA and ACB are duals when treatment C is fixed in the second period.

Thus,

$$\sum_{k' \in dual sequences of k} \mathbf{X}_{k'}^{\tau} = (l_k - 1)! C_{t-1}^{l_k - 1} \mathbf{1}_{[p]} \otimes \mathbf{1}_{[t]}^{T},$$

Hence, we have

$$\begin{aligned}
\overline{\mathbf{X}}^{\tau} &= \frac{1}{N} \overline{\mathbf{V}^{-1}}^{-1} \sum_{km} n_{km} \mathbf{V}_m^{-1} \mathbf{X}_k^{\tau} \\
&= \frac{1}{N} \overline{\mathbf{V}^{-1}}^{-1} \sum_{m} \mathbf{V}_m^{-1} \sum_{k} n_{km} \mathbf{X}_k^{\tau} \\
&= \frac{1}{N} \overline{\mathbf{V}^{-1}}^{-1} \sum_{m} \mathbf{V}_m^{-1} \sum_{k'} n_{k'm} \sum_{k \in duals of k'} \mathbf{X}_k^{\tau} \\
&= \frac{1}{N} \overline{\mathbf{V}^{-1}}^{-1} \sum_{m} \mathbf{V}_m^{-1} \sum_{k'} n_{k'm} (l_{k'} - 1)! C_{t-1}^{l_{k'} - 1} \mathbf{1}_{[p]} \otimes \mathbf{1}_{[t]}^{T} \\
&= \frac{1}{N} \overline{\mathbf{V}^{-1}}^{-1} \sum_{m} \mathbf{V}_m^{-1} \sum_{k'} n_{k'm} (l_{k'})! C_{t}^{l_{k'}} \mathbf{1}_{[p]} \otimes \mathbf{1}_{[t]}^{T} / t \\
&= \frac{1}{N} \overline{\mathbf{V}^{-1}}^{-1} \sum_{m} \mathbf{V}_m^{-1} \sum_{k} n_{km} \mathbf{1}_{[p]} \otimes \mathbf{1}_{[t]}^{T} / t \\
&= \mathbf{1}_{[p]} \otimes \mathbf{1}_{[t]}^{T} / t,
\end{aligned}$$

where k' runs through sequences which are dual to only themselves. Similarly, $\overline{\mathbf{X}}^{\gamma} = \mathbf{1}_{[p]}^{*} \otimes \mathbf{1}_{[t]}^{T} / t$.

Now, we only need to prove that $\mathbf{C} = \sum_{km} n_{km} (\mathbf{X}_k^{\tau})^{T} \mathbf{V}_m^{-1} \mathbf{X}_k^{\tau} = (a - b) \mathbf{I}_{[t]} + b \mathbf{1}_{[t]} \mathbf{1}_{[t]}^{T}$, where a and b are two constants. For any sequence k, the design matrices of its duals can be generated by $\mathbf{X}_k^{\tau} \mathbf{A}$ and $\mathbf{X}_k^{\gamma} \mathbf{A}$, where $\mathbf{A}$ is obtained by permuting the columns $(\{\mathbf{e}_i\})$of $\mathbf{I}_{[l_k]}$. Here, $\mathbf{e}_i$ is the vector of 0's with the i^{th} element replaced by 1. There are totally $l_k! C_t^{l_k}$ such $\mathbf{A}$. For the dual

sequences k^*, $(\mathbf{X}_{k^*}^{\tau})^T\mathbf{V}_m^{-1}\mathbf{X}_{k^*}^{\tau} = \mathbf{A}^T(\mathbf{X}_k^{\tau})^T\mathbf{V}_m^{-1}\mathbf{X}_k^{\tau}\mathbf{A}$.

Note that among all the $\mathbf{A}$'s, $\mathbf{e}_i$, $i = 1, \ldots, l_k$ appear with equal frequency in any column. Thus, by the definition of dual-balanced symmetric design, we have that the diagonal elements of $\mathbf{C}$ are all equal. Further, note that, if any two columns are selected, the resulting pairs $(\mathbf{e}_i, \mathbf{e}_j)$ do not depend on which columns are selected. Hence, the non-diagonal elements of $\mathbf{C}$ are all equal to each other, i.e., $\mathbf{C} = (a-b)\mathbf{I}_{[t]} + b\mathbf{1}_{[t]}\mathbf{1}_{[t]}^T$. By Lemma 3.2.2, we have that $\mathbf{I}_{11} = c(t\mathbf{I}_{[t]} - \mathbf{1}_{[t]}\mathbf{1}_{[t]}^T)$. Similar results can be obtained for $\mathbf{I}_{12}$ and $\mathbf{I}_{22}$. Then by Lemma 3.2.1 and Definition 3.2.2, we have 3.2.2.

The elements are obtained as $a_{11}^{km} = tr[(\mathbf{X}_k^{\tau} - \overline{\mathbf{X}}_k^{\tau})^T\mathbf{V}_m^{-1}(\mathbf{X}_k^{\tau} - \overline{\mathbf{X}}_k^{\tau})] = tr(\mathbf{X}_k^{\tau T}\mathbf{V}_m^{-1}\mathbf{X}_k^{\tau}) - tr(\overline{\mathbf{X}}_k^{\tau T}\overline{\mathbf{V}^{-1}\mathbf{X}_k^{\tau}})$, $a_{12}^{km} = tr(\mathbf{X}_k^{\tau T}\mathbf{V}_m^{-1}\mathbf{X}_k^{\gamma}) - tr(\overline{\mathbf{X}}^{\tau T}\overline{\mathbf{V}^{-1}\mathbf{X}^{\gamma}})$ and $a_{22}^{km} = tr(\mathbf{X}_k^{\gamma T}\mathbf{V}_m^{-1}\mathbf{X}_k^{\gamma}) - tr(\overline{\mathbf{X}}^{\gamma T}\overline{\mathbf{V}^{-1}\mathbf{X}^{\gamma}})$, and the results follow by noting that

$$\begin{aligned}
tr(\overline{\mathbf{X}}_k^{\tau T}\overline{\mathbf{V}^{-1}\mathbf{X}_k^{\tau}}) &= \frac{1}{t^2}tr(\mathbf{1}_{[p]} \otimes \mathbf{1}_{[t]}^T\overline{\mathbf{V}^{-1}}\mathbf{1}_{[p]}^T \otimes \mathbf{1}_{[t]}) \\
&= \frac{1}{t^2}tr(\mathbf{1}_{[t]}^T \otimes \mathbf{1}_{[p]}\overline{\mathbf{V}^{-1}}\mathbf{1}_{[t]} \otimes \mathbf{1}_{[p]}^T) \\
&= \frac{1}{t^2}tr((\mathbf{1}_{[t]}\mathbf{1}_{[t]}^T) \otimes \mathbf{1}_{[p]}^T\overline{\mathbf{V}^{-1}}\mathbf{1}_{[p]}) \\
&= \frac{1}{t}\mathbf{1}_{[p]}^T\overline{\mathbf{V}^{-1}}\mathbf{1}_{[p]} \\
&= \frac{1}{t}v_1
\end{aligned}$$

and that

$$\begin{aligned} tr(\overline{\mathbf{X}_k^{\gamma}}^T\overline{\mathbf{V}^{-1}\mathbf{X}_k^{\gamma}}) &= \frac{1}{t^2}tr(\mathbf{1}_{[p]} \otimes \mathbf{1}_{[t]}^T\overline{\mathbf{V}^{-1}}\mathbf{1}_{[p]}^{*\,T} \otimes \mathbf{1}_{[t]}) \\ &= \frac{1}{t}\mathbf{1}_{[p]}^T\overline{\mathbf{V}^{-1}}\mathbf{1}_{[p]}^{*} \\ &= \frac{1}{t}v_2, \end{aligned}$$

$$\begin{aligned} tr(\overline{\mathbf{X}_k^{\gamma}}^T\overline{\mathbf{V}^{-1}\mathbf{X}_k^{\gamma}}) &= \frac{1}{t^2}tr(\mathbf{1}_{[p]}^{*} \otimes \mathbf{1}_{[t]}^T\overline{\mathbf{V}^{-1}}\mathbf{1}_{[p]}^{*\,T} \otimes \mathbf{1}_{[t]}) \\ &= \frac{1}{t}\mathbf{1}_{[p]}^{*\,T}\overline{\mathbf{V}^{-1}}\mathbf{1}_{[p]}^{*} \\ &= \frac{1}{t}v_3. \end{aligned}$$

$\square$

Corollary 3.2.1 *The information matrix for the treatment effects is*

$$\mathbf{I}(\boldsymbol{\tau}) = (a_{11} - a_{12}^2/a_{22})(t\mathbf{I}_{[t]} - \mathbf{1}_{[t]}\mathbf{1}_{[t]}^T)/t(t-1) \qquad (3.2.3)$$

Corollary 3.2.2 *When the subject effects are fixed, a_{11}^{km}, a_{12}^{km} and a_{22}^{km} in Theorem (3.2.1) can be simplified to $a_{11}^{km} = \sum_{i,i'=1}^{p} \delta_k^{ii'}(\mathbf{V}_m^{-1})_{i,i'}$, $a_{12}^{km} = \sum_{i=1}^{p}\sum_{i'=1}^{p-1} \delta_k^{ii'}$ $(\mathbf{V}_m^{-1})_{i,i'+1}$ and $a_{22}^{km} = \sum_{i=1}^{p-1}\sum_{i'=1}^{p-1} \delta_k^{ii'}(\mathbf{V}_m^{-1})_{i+1,i'+1} - v_4/t$, where $v_4 = \sum_{km} n_{km}$ $(\mathbf{V}_m^{-1})_{1,1}/N$. When $M = 1$, components $a_{ll'}^{km}$, for $l, l' = 1, 2$, reduce to those in Kushner (2000).*

PROOF. Note that, when subject effects are fixed, the matrix $\mathbf{V}_{jk}^{-1} = \mathbf{\Sigma}_{jk}^{-1} - \mathbf{\Sigma}_{jk}^{-1}\mathbf{1}_{[p]}\mathbf{1}_{[p]}^T\mathbf{\Sigma}_{jk}^{-1}/\mathbf{1}_{[p]}^T\mathbf{\Sigma}_{jk}^{-1}\mathbf{1}_{[p]}$ is spanned by $\mathbf{1}_{[p]}$, i.e. $\mathbf{V}_{jk}\mathbf{1}_{[p]} = \mathbf{0}$. Thus Corollary 3.2.2 follows.

□

3.3 Allocation rules for symmetric designs

We shall restrict our attention in this section to balanced symmetric designs, as dual-balanced designs are found to be optimal in the literature. When the inferential interest is in τ, we focus on the information matrix in (3.2.3).

Note that, since the information for $\boldsymbol{\tau}$ depends on $\boldsymbol{n}$ only through the function $f(\boldsymbol{n}) = a_{11} - a_{12}^2/a_{22}$, the increment of the information from new subjects $\boldsymbol{\Delta n} = (\Delta n_{11}, \ldots, \Delta n_{1M}, \ldots, \Delta n_{s1}, \ldots, \Delta n_{sM})^T$ is determined by the increment of f, which is $\Delta f = f(\boldsymbol{n} + \boldsymbol{\Delta n}) - f(\boldsymbol{n})$. Intuitively, programs can be used to sift out the maximum of $f(\boldsymbol{n} + \boldsymbol{\Delta n})$ by thoroughly sorting over all possible $\boldsymbol{\Delta n}$ satisfying $\mathbf{1}^T\boldsymbol{\Delta n} = \Delta N$, where ΔN is the total number of new subjects at each allocation step. However, such programs can be very time consuming and costly. The level of cost depends on the size of ΔN. As in Kushner (2000), we approximate Δf using the first order of Taylor's Expansion

$$\Delta f = [\frac{\partial f}{\partial \boldsymbol{n}}]^T \boldsymbol{\Delta n},$$

where it can be shown that

$$\frac{\partial f}{\partial \boldsymbol{n}} = (q_{11}(c_0; \mathbf{V}_1), \ldots, q_{1M}(c_0; \mathbf{V}_M), \ldots, q_{s1}(c_0; \mathbf{V}_1), \ldots, q_{sM}(c_0; \mathbf{V}_M))^T$$

with each component

$$\begin{aligned} q_{km}(c; \mathbf{V}_m) &= tr\{[\mathbf{X}_k^{\tau} - \overline{\mathbf{X}}^{\tau} + c(\mathbf{X}_k^{\gamma} - \overline{\mathbf{X}}^{\gamma})]^T \mathbf{V}_m^{-1} [\mathbf{X}_k^{\tau} - \overline{\mathbf{X}}^{\tau} + c(\mathbf{X}_k^{\gamma} - \overline{\mathbf{X}}^{\gamma})]\} \\ &= a_{11}^{km} + 2ca_{12}^{km} + c^2 a_{22}^{km} \end{aligned}$$

and $c_0 = -a_{12}/a_{22}$.

Thus,

$$\Delta f = \sum_{k,m} \Delta n_{km} \times q_{km}(c_0; \mathbf{V}_m). \tag{3.3.1}$$

We used the following allocation rules to optimize the experiment at each allocation step.

Allocation Rule 1: *The increment of information is maximized by assigning all new subjects associated with covariance matrix* $\mathbf{V}_m$ *to sequence* K *and its dual sequences, such that*

$$q_{Km}(c_0; \mathbf{V}_m) = \max_k q_{km}(c_0; \mathbf{V}_m) \tag{3.3.2}$$

subject to

$$s_K \times \Delta n_{Km} = \Delta N, \tag{3.3.3}$$

where Δn_{Km} *is an integer and* s_K *is the number of dual sequences to sequence* K*, including itself. Hence,* $\Delta n_{km} = 0$ *for the rest sequences that are not the dual sequence of sequence* K*.*

This rule reduces to the allocation rule in Kushner (2000) when a homogeneous covariance matrix is assumed for all subjects.

If we can further assume that $\Delta n_{km} = \Delta n_{k^* m} = 1$, then a more accurate approximation for Δf can be obtained via a second order-expansion, as

$$\Delta f = [\frac{\partial f}{\partial \boldsymbol{n}}]^T \boldsymbol{\Delta n} + \frac{1}{2} \boldsymbol{\Delta n}^T [\frac{\partial^2 f}{\partial \boldsymbol{n}^2}] \boldsymbol{\Delta n},$$

where

$$\frac{\partial^2 f}{\partial \boldsymbol{n}^2} = -\frac{1}{2a_{22}} \frac{\partial q_{km}}{\partial c}|_{c=c_0} \frac{\partial q_{k'm'}}{\partial c}|_{c=c_0}.$$

Thus, we can refine Δf as

$$\Delta f = \sum_{k,m} \Delta n_{km} \times q_{km}(c_0; \mathbf{V}_m) - \frac{1}{4a_{22}} [\sum_{k,m} \Delta n_{km} \times \frac{\partial q_{km}}{\partial c}|_{c=c_0}]^2 \tag{3.3.4}$$

so that the following allocation rule is possible.

Allocation Rule 2: *When each new subject is to be allocated to a sequence, the allocation rule given in (3.3.2) is refined to assign all new subjects associated* *with covariance matrix* $\mathbf{V}_m$ *to sequence* K *and its dual sequences by choosing* $\hat{K}$ *such that*

$$q_{Km}(c_0;\mathbf{V}_m) - \frac{1}{4a_{22}}[\frac{\partial q_{Km}}{\partial c}|_{c=c_0}]^2 = \max_k\{q_{km}(c_0;\mathbf{V}_m) - \frac{1}{4a_{22}}[\frac{\partial q_{km}}{\partial c}|_{c=c_0}]^2\} \tag{3.3.5}$$

with

$$\frac{\partial q_{km}}{\partial c}|_{c=c_0} = 2(a_{12}^{km} + c_0 a_{22}^{km}).$$

Here, we assume

$$\Delta n_{Km} = \Delta n_{K^*m} = 1,$$

$$\Delta N = s_K, \tag{3.3.6}$$

where K^* *is the dual sequence of* K.

Note that it may be impossible to assume *a priori* which covariance matrix is associated with new subjects. If the actual covariance matrix is not $\mathbf{V}_m$, as assumed in allocation rules 1 and 2, then the two allocation rules do not necessarily maximize the increment of information on τ. To deal with such situations, we adjust the aforementioned allocation rules in the following manner.

First, we define Δf^* as the true increment of information when the covariance matrix associated with new subjects is correctly specified. Depending on which of $\mathbf{V}_m$, $m = 1, \ldots, M$, is the true covariance matrix, Δf^* takes values $\Delta f(\mathbf{V}_1), \ldots, \Delta f(\mathbf{V}_M)$, where $\Delta f(\mathbf{V}_m)$ is the increment information (3.3.1) with Δn_{km} and q_{km} substituted by (3.3.2) and (3.3.3) in Allocation Rule 1, or the information (3.3.4) with Δn_{km} and q_{km} substituted by (3.3.6) and (3.3.5) in Allocation Rule 2. The probability of Δf^* taking value $\Delta f(\mathbf{V}_m)$, denoted by p_m, is the same as the probability that the new subjects have covariance matrix $\mathbf{V}_m$. Here, $\{p_1, \ldots, p_M\}$ are known or could be obtained from a pilot study.

Allocation Rule 3: *The subjects are assigned to sequence K, such that the loss of increment information due to an incorrect assumption about the associated variance is minimized.*

$$\begin{aligned} & \min_m E_{\Delta f^*}[(\Delta f(\mathbf{V}_m) - \Delta f^*)^2] && (3.3.7) \\ = \; & \min_m \sum_{m'=1}^{M} p_{m'}[\Delta f(\mathbf{V}_m) - \Delta f(\mathbf{V}_{m'})]^2. \end{aligned}$$

The value of K depends on $\mathbf{V}_m$ as stated in allocation rules 1 and 2. If the minimum of (3.3.7) is obtained at m^*, then new subjects are assigned to sequence K and its duals, as obtained from (3.3.2) or (3.3.5) using $\mathbf{V}_{m^*}$.

3.4 Implementation of the rules

A naive procedure for implementing the allocation rules is to update the estimates for $\{\mathbf{V}_m\}$ and β adaptively once the assigned subjects complete the experiment. Then, applying the updated estimates to the allocation rules, we continue to assign all subjects to treatment sequences. Specifically, the procedure consists of the following steps.

Step 1. Start the experiment with an initial N_0 subjects using the optimal or "nearly" optimal design suggested in the literature. Compute the maximum likelihood estimates $\hat{\mathbf{V}}_m(\boldsymbol{n})$, $m = 1, \ldots, M$ and $\hat{\boldsymbol{\beta}}(\boldsymbol{n})$ with vector $\boldsymbol{n}$ representing the initial design, iteratively, by

$$\hat{\mathbf{V}}_m(\boldsymbol{n}) = \sum_{k=1}^{s} \sum_{j=1}^{n_{km}} (\mathbf{y}_{jk} - \mathbf{X}_k \hat{\boldsymbol{\beta}}(\boldsymbol{n}))(\mathbf{y}_{jk} - \mathbf{X}_k \hat{\boldsymbol{\beta}}(\boldsymbol{n}))^T, \tag{3.4.1}$$

and

$$\hat{\boldsymbol{\beta}}(\boldsymbol{n}) = (\sum_{km} \mathbf{X}_k^T \hat{\mathbf{V}}_m(\boldsymbol{n})^{-1} \mathbf{X}_k)^{-1} \sum_{k=1}^{s} \sum_{j=1}^{n_{km}} \mathbf{X}_k^T \hat{\mathbf{V}}_{jk}(\boldsymbol{n})^{-1} \mathbf{y}_{jk}. \tag{3.4.2}$$

Step 2. Assign new subjects to treatment sequences using allocation rules (3.3.2 or 3.3.5 along with 3.3.7) with $\mathbf{V}_m$ substituted by its estimates as in Step 1.

Step 3. Update the estimates $\hat{\mathbf{V}}_m(\boldsymbol{n}+\boldsymbol{\Delta n})$ and $\hat{\boldsymbol{\beta}}(\boldsymbol{n}+\boldsymbol{\Delta n})$ using the subjects in Steps 1 and 2.

Step 4. Repeat Steps 1-3 until all subjects have been allocated.

3.5 Efficiency of adaptive designs

To assess the efficiency of the adaptive designs, we compare its matrix of mean squared error

$$\mathbf{MSE} = E[(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau})(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau})^T]$$

with that of the "optimal" designs in a given situation. In the simulation study, the **MSE** is estimated by

$$\hat{\mathbf{MSE}} = \sum_{b=1}^{B}(\hat{\boldsymbol{\tau}}^{(b)} - \boldsymbol{\tau})(\hat{\boldsymbol{\tau}}^{(b)} - \boldsymbol{\tau})^T/B,$$

where $\hat{\boldsymbol{\tau}}^{(b)}$ is the MLE obtained in the b^{th} simulation run, $B = 1000$. We denote $\mathbf{MSE}_1$ as the matrix of mean squared error for the proposed adaptive design and $\mathbf{MSE}_0$ for the adaptive design under the variance-covariance homogeneity assumption. Based on A-, D- or E- optimality criteria (Kiefer 1975), we defined the efficiency of the proposed design as

$$eff_A = \frac{tr(\mathbf{MSE}_1)}{tr(\mathbf{MSE}_0)}$$

$$eff_D = \frac{|\mathbf{MSE}_1|}{|\mathbf{MSE}_0|}$$

$$eff_E = \frac{\max eigenvalue(\mathbf{MSE}_1)}{\max eigenvalue(\mathbf{MSE}_0)}$$

For a singular matrix, the determinant is defined as the product of non-zero eigenvalues. The efficiency comparison was conducted by simulation for two-period two-treatment; three-period two-treatment; and three-period three-treatment designs for $M = 1, 2, 3$. The case of $M = 1$ is the case of assuming homogeneity. For two-treatment designs, since the information matrix (3.2.3) becomes

$$\mathbf{I}(\boldsymbol{\tau}) = a \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix},$$

the trace, the maximum eigenvalue and the determinant of $\mathbf{I}(\boldsymbol{\tau})$ are all equal to $2a$. Thus, all three criteria are equivalent.

3.5.1 $t = 2$ and $p = 2$

We applied the proposed allocation rules to the two-treatment two-period design, consisting of four sequences, AA, AB, BA and BB. For $M = 2$, we assumed that all the subjects were associated with two different covariance matrices

$$\mathbf{V}_1 = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

and

$$\mathbf{V}_2 = \begin{bmatrix} 1 & c_1\rho \\ c_1\rho & c_1^2 \end{bmatrix}$$

with $\rho = .3, .7$, $c_1 = 2$ and the proportion of subjects associated with V_2, $P = \sum_k n_{k2}/N$, controlled to be 20% or 50%. For $M = 3$, the three different covariance matrices are $\mathbf{V}_1$, $\mathbf{V}_2$ with $\rho = .3$ and $c_1 = 2$, and $\mathbf{V}_3$ is obtained from $\mathbf{V}_2$ with $c_1 = 2$ and $\rho = .7$. The proportions of subjects associated with these three covariance matrices were 50%, 30% and 20%, respectively. Initially, four subjects were assigned to sequences AB and BA, respectively. The experiment included 100 subjects in total.

First, we show the performance of the adaptive design ($M = 1$) compared to the fixed optimal design using the covariance matrix $\mathbf{V}_1$ with $\rho = .3$. The fixed optimal design for this case is the design with an equal number of subjects assigned to sequences AA, AB and their duals. Figure 3.1 demonstrates the relative efficiency as the ratio of the MSE for the fixed design to that for the adaptive design. The variance for τ when $\rho = .3$ is $2.06/N$. The adaptive design was as efficient as the optimal design, when $\mathbf{V}_1$ is assumed known before the experiment. It has an efficiency over 95%, when the sample is at least 10. The final efficiency when $N = 100$ is 97%.

We now focus on situations where the fixed optimal design is not as easily identifiable as in the above example because there is no prior information on the covariance matrix. In the following discussions, we shall compare only the adaptive designs to assess the merits or disadvantages of incorporating heterogeneity into design construction.

Figure 3.2 is a comparison of the adaptive designs when two different covariance matrices associated with the subjects and the two periods were weakly correlated with $\rho = .3$ against the design that assumed homogeneity. When the sample size is relatively small, say less than 30, it shows that incorporating heterogeneity leads to less efficient designs. The reason may be that there are not yet enough subjects in the experiment to be able to distinguish between the two different covariance matrices. However, as the sample size increases, it becomes evident that incorporating heterogeneity leads to high-efficiency designs, especially when the two covariance matrices are notably different from each other and equally likely ($P = 50\%$).

Figure 3.3 illustrates the relative efficiency of our two-treatment two-period designs when $M = 2$ and $\rho = .7$. The findings are similar to those described above. Comparing the curves in Figure 3.2 and Figure 3.3, we find that our strategy improves as ρ increases with P and c_1 fixed. At $N = 100$, the efficiency when $c_1 = 2$ and $P = 50\%$ is about 67% when $\rho = .7$, about 75% when $\rho = .3$. When $c_1 = 2, P = 20\%$ and $\rho = .7$, the efficiency is 77%, and, it is about 87% when $\rho = .3$. This indicates that incorporating heterogeneity improves the efficiency by almost one and a half times.

Figure 3.4 demonstrates that more subjects are needed to distinguish the different covariance matrices when M increases to 3. Our approach becomes superior and the relative efficiency attains 70% when the sample size increases to 100.

3.5.2 $t = 2$ and $p = 3$

For two-treatment three-period designs, the proposed allocation rules were applied to assign subjects to eight possible sequences: AAA, AAB, ABA, ABB and their duals. For $M = 2$, we assumed all subjects were associated with two different covariance matrices,

$$\mathbf{V}_1 = \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{bmatrix}$$

and

$$\mathbf{V}_2 = \begin{bmatrix} 1 & c_2\rho_1 & c_1\rho_2 \\ c_2\rho_1 & c_2^2 & c_1\rho_3 \\ c_1\rho_2 & c_1\rho_3 & c_1^2 \end{bmatrix}$$

with $\rho_1 = \rho_2 = \rho_3 = .3, .7$ or $\rho_1 = \rho_3 = \sqrt{\rho_2}$ and $c_1 = 1, c_2 = 2$ or $c_1 = c_2 = 2$. The proportion of subjects associated with $\mathbf{V}_2$ is 50% in our simulation. For $M = 3$, we considered three different covariance matrices: $\mathbf{V}_1$, $\mathbf{V}_2$ with $c_1 = 1$ and $c_2 = 2$, and $\mathbf{V}_3$ obtained as in $\mathbf{V}_2$ with $c_1 = c_2 = 2$. The proportions of subjects associated with these three covariance matrices were 50%, 30% and 20%, respectively. Note that $\mathbf{V}_1$ with $\rho_1 = \rho_2 = \rho_3$ has an equicorrelated covariance structure, and $\mathbf{V}_1$ with $\rho_1 = \rho_3 = \sqrt{\rho_2}$ is the first-order autoregressive covariance structure. Initially, six subjects were assigned to sequences ABB and BAA, respectively, according to the optimal design suggested by Kershner and Federer (1981) and Laska et al. (1983) for equicorrelated covariance structure.

Figures 3.5-3.8 summarize the simulation results for the comparison of relative efficiencies of the adaptive designs for two-treatment three-period designs. Similar to what we observed with two-treatment two-period designs, the proposed design strategy was notably superior. In the case of $\rho_1 = \rho_3 = \sqrt{\rho_2}$, the disparity of the covariance matrices did not affect the performance of the proposed design as much as in the case of $\rho_1 = \rho_2 = \rho_3$. The final efficiency was 81% when $c_1 = c_2 = 2$, and about 90% when $c_1 = 1, c_2 = 2$ and all $\rho = .3$. With the correlation coefficients increasing to .7, the efficiencies became 73% and 85%, respectively. In the case of autoregressive errors, the final efficiency is similar to that for equal correlation coefficients, when $c_1 = 1$ and $c_2 = 2$.

Figures 3.9-3.12 demonstrate the performance of our approach when $M = 3$. The relative efficiency depends on the degree of disparity of the covariance matrices. The final efficiency at $N = 100$ is around 85%, and 75% when the correlation coefficient is .3 and .7, respectively, for an equicorrelated covariance structure. In the case of autoregressive error, the efficiency increased to 90% for $\rho_1 = .3$ and 80% when $\rho_1 = .7$.

3.5.3 $t = 3$ and $p = 3$

The same situations as in 3.4.2 were considered with two subjects initially in each of the sequences ABB, ACC, BAA, BCC, CAA and CBB.

Figures 3.11-3.18 demonstrate the results of comparisons of the relative efficiency of adaptive designs with $M = 2$ against that with $M = 1$. The efficiency curves based on the three optimal criteria (A-, D- and E-optimal) are shown for each case. The efficiency curves based on A- and E- optimal criteria are similar, while the one based on D-optimal criteria is about 8-15% lower. The final efficiency based on D-optimal criteria for the case of $c_1 = c_2 = 2$ is 73%, which is 12% lower than that for the case of $c_1 = 1, c_2 = 2$, when all correlation coefficients are equal to .3. When correlation coefficients increase to .7, the final efficiency based on D-optimal criteria is 14% lower, as compared with 73% as c_1 increases from 1 to 2.

Similar to the two-treatment cases (not shown), an increasing number of different covariance matrices requires more subjects to be effective. This is even more important when ρ is not very high.

3.6 Conclusion

Recognizing that the study subjects in an experiment can be heterogeneous in their responses to the treatments given, and that they usually enter the experiment sequentially, we developed a methodology for constructing adaptive designs suitable for these situations. Our strategy was to minimize the loss function of the increment of information added by each new set of subjects. The resulting designs were shown to be more efficient than those that do not account for heterogeneity, and robust to the assumptions required in traditional design construction.

The superiority of the proposed design strategy becomes more evident when covariance matrices differ greatly. When the number of treatments is more than two, the efficiency of our strategy depends on the optimality criteria. The simulation results showed that A-optimal and E-optimal criteria led to similar efficiency curves for all the models we considered. Efficiency curves based on the D-optimal criterion behave quite differently from the other two. It appears that D-optimal criterion is more sensitive to variance heterogeneity and requires larger sample sizes than the other two criteria for the proposed adaptive designs to be effective.

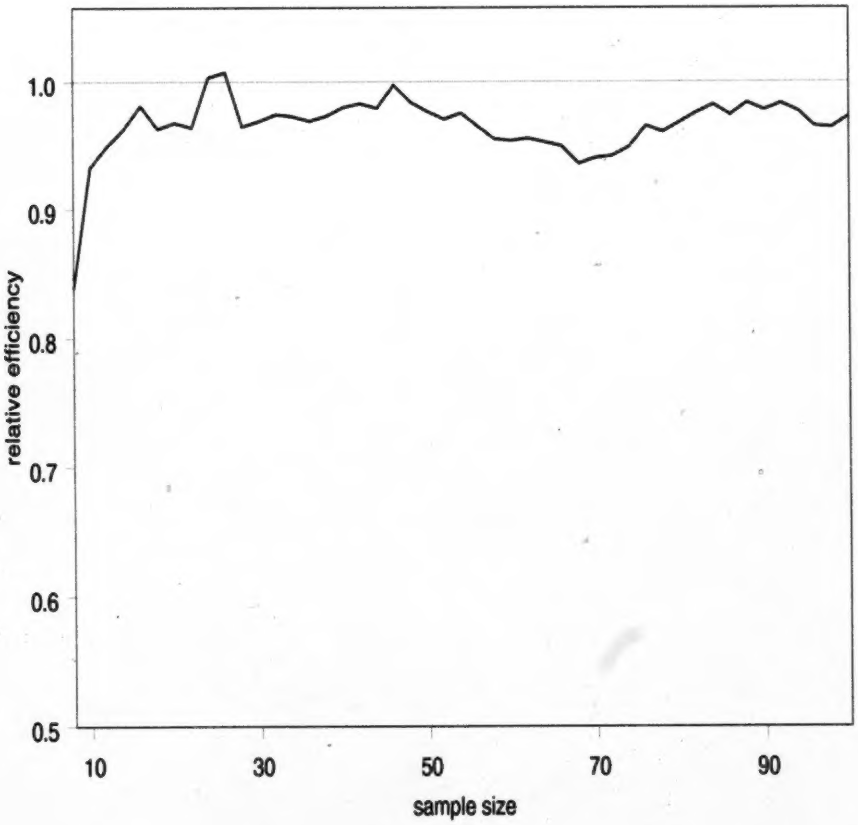

Figure 3.1: Relative efficiency curves for adaptive two-treatment two-period designs with covariance matrix $\mathbf{V}_1$ in Section 3.5.1 with $\rho = .3$ for $M = 1$. The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., close to the line) relative to the fixed optimal design.

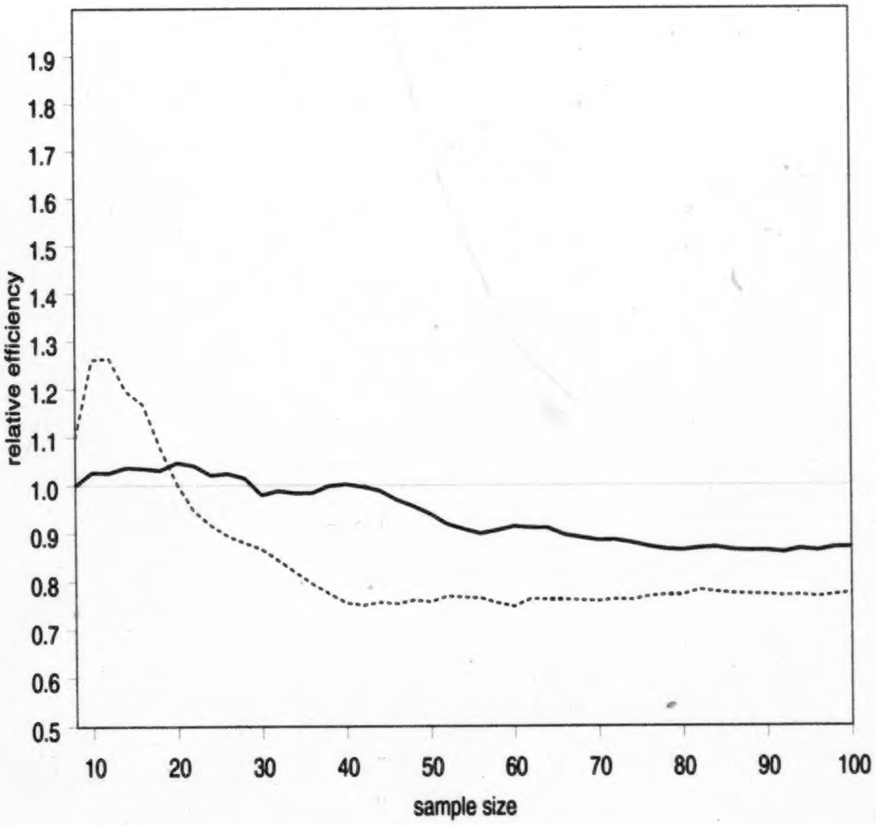

Figure 3.2: Relative efficiency curves for adaptive two-treatment two-period designs with correlation coefficient $\rho = .3$ for $M = 2$ for $P = 20\%$ and $c_1 = 2$ (the solid line) and $P = 50\%$ and $c_1 = 2$ (the dashed line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$).

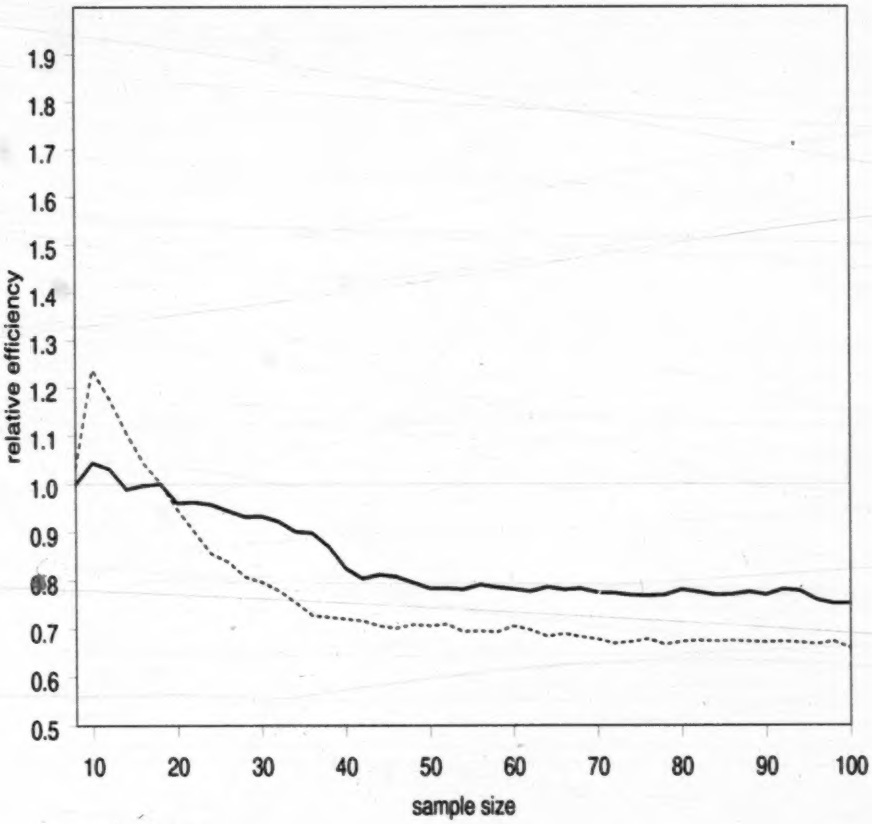

Figure 3.3: Relative efficiency curves for adaptive two-treatment two-period designs with correlation coefficient $\rho = .7$ for $M = 2$ for $P = 20\%$ and $c_1 = 2$ (the solid line) and $P = 50\%$ and $c_1 = 2$ (the dashed line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$).

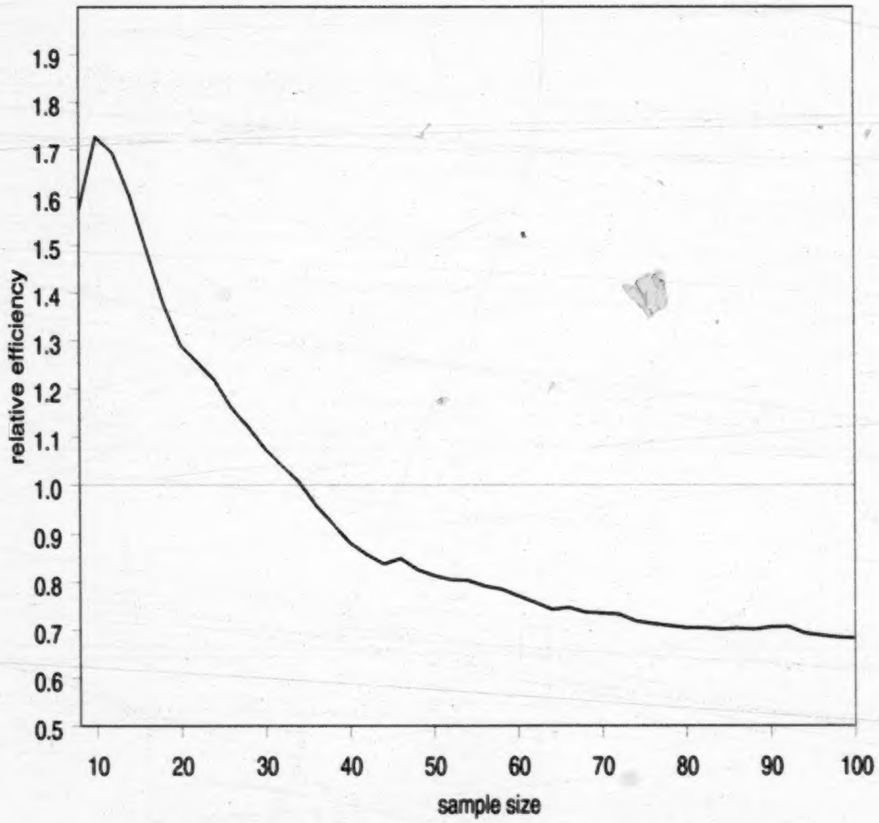

Figure 3.4: Relative efficiency curves for adaptive two-treatment two-period designs for $M = 3$. The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$).

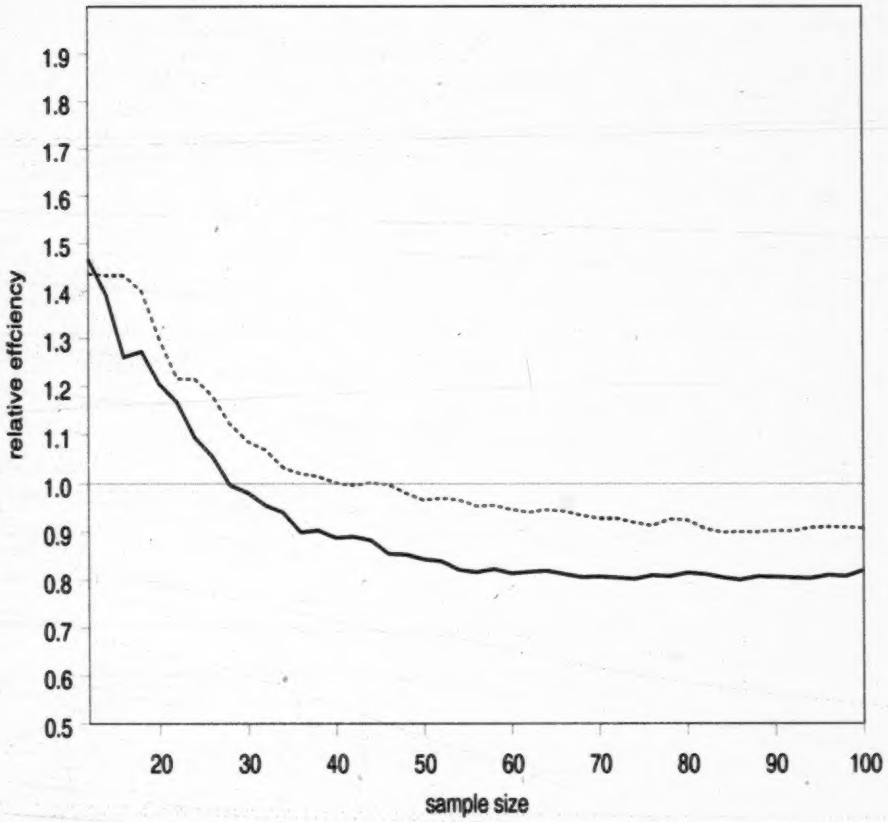

Figure 3.5: Relative efficiency curves for adaptive two-treatment three-period designs when $M = 2$ and $\rho_1 = \rho_2 = \rho_3 = .3$ for $c_1 = c_2 = 2$ (the solid line) and $c_1 = 1$ and $c_2 = 2$ (the dashed line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$).

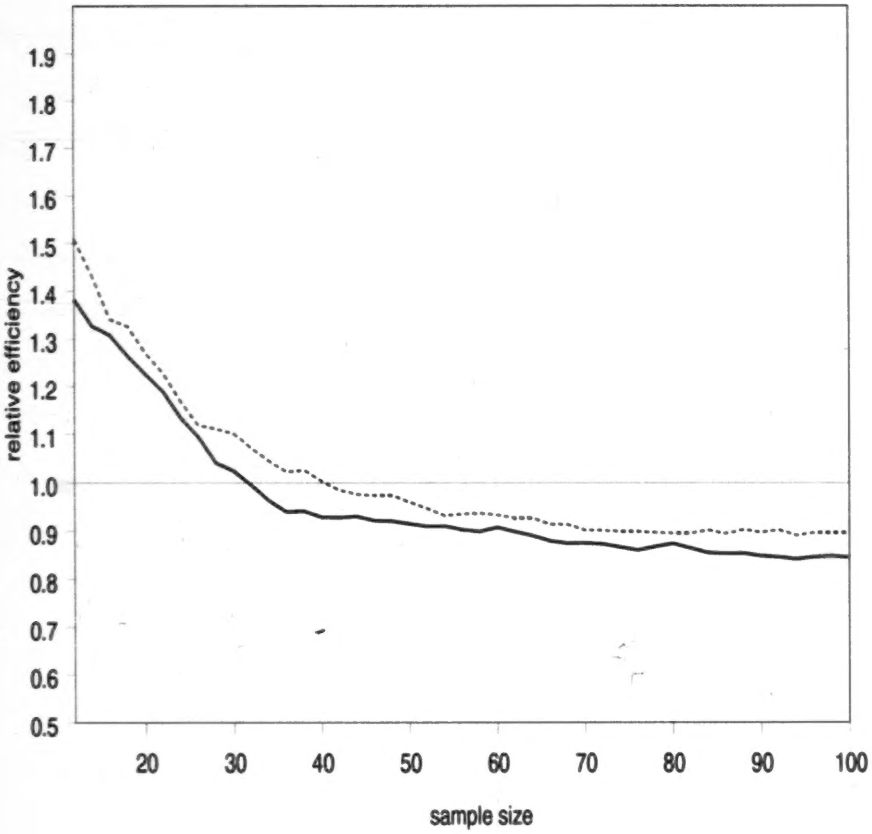

Figure 3.6: Relative efficiency curves for adaptive two-treatment three-period designs when $M = 2$ and $\rho_1 = \rho_3 = \sqrt{\rho_2} = .3$ for $c_1 = c_2 = 2$ (the solid line) and $c_1 = 1$ and $c_2 = 2$ (the dashed line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$).

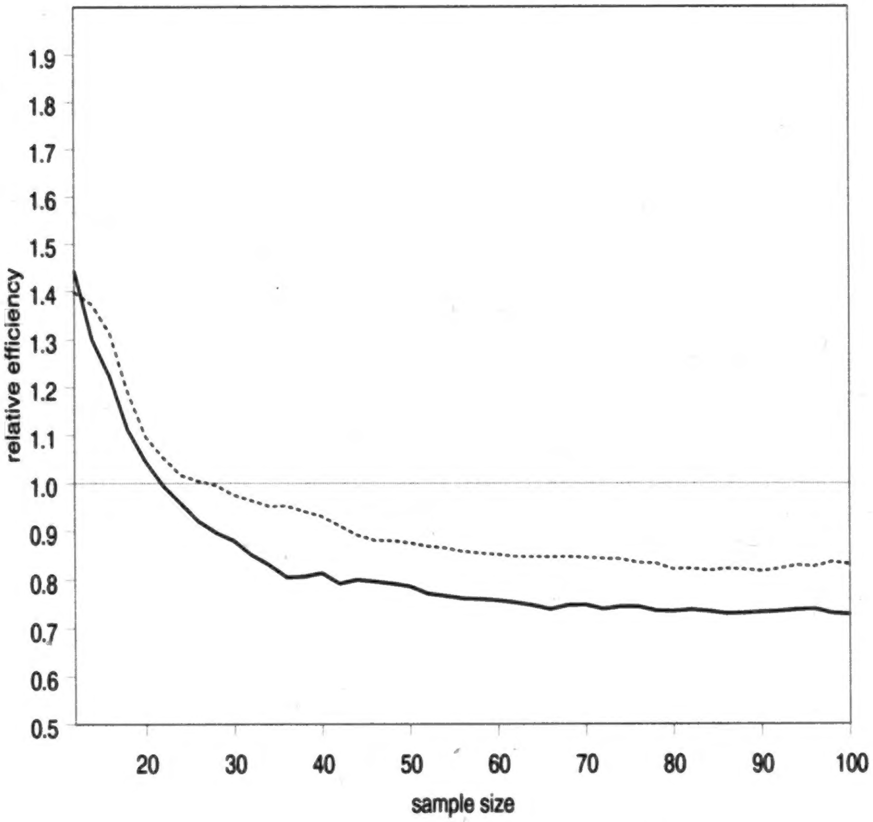

Figure 3.7: Relative efficiency curves for adaptive two-treatment three-period designs when $M = 2$ and $\rho_1 = \rho_2 = \rho_3 = .7$ for $c_1 = c_2 = 2$ (the solid line) and $c_1 = 1$ and $c_2 = 2$ (the dashed line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$).

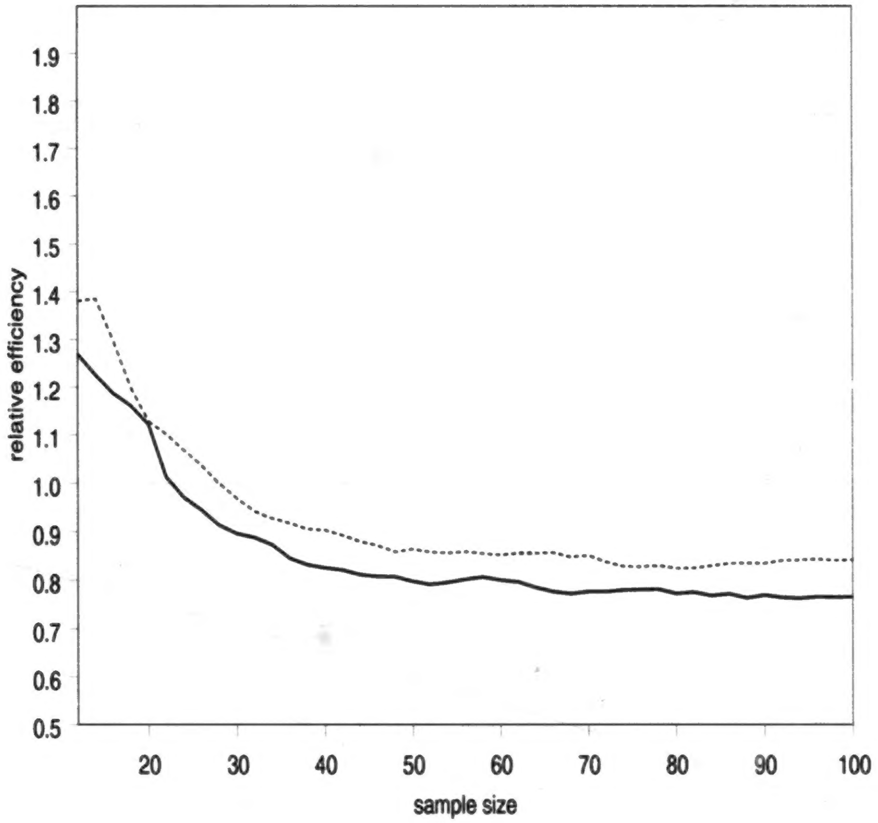

Figure 3.8: Relative efficiency curves for adaptive two-treatment three-period designs when $M = 2$ and $\rho_1 = \rho_3 = \sqrt{\rho_2} = .7$ for $c_1 = c_2 = 2$ (the solid line) and $c_1 = 1$ and $c_2 = 2$ (the dashed line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$).

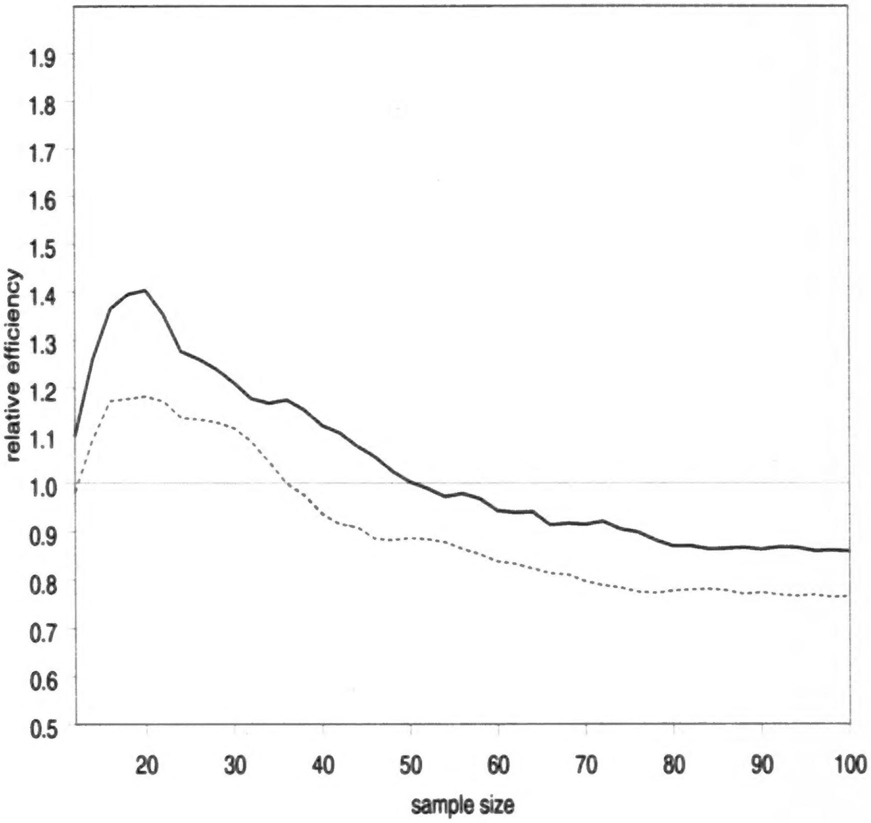

Figure 3.9: Relative efficiency curves for adaptive two-treatment three-period designs when $M = 3$ and $\rho_1 = \rho_2 = \rho_3 = .3$ (the solid line), and $\rho_1 = \rho_2 = \rho_3 = .7$ (the dashed line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$).

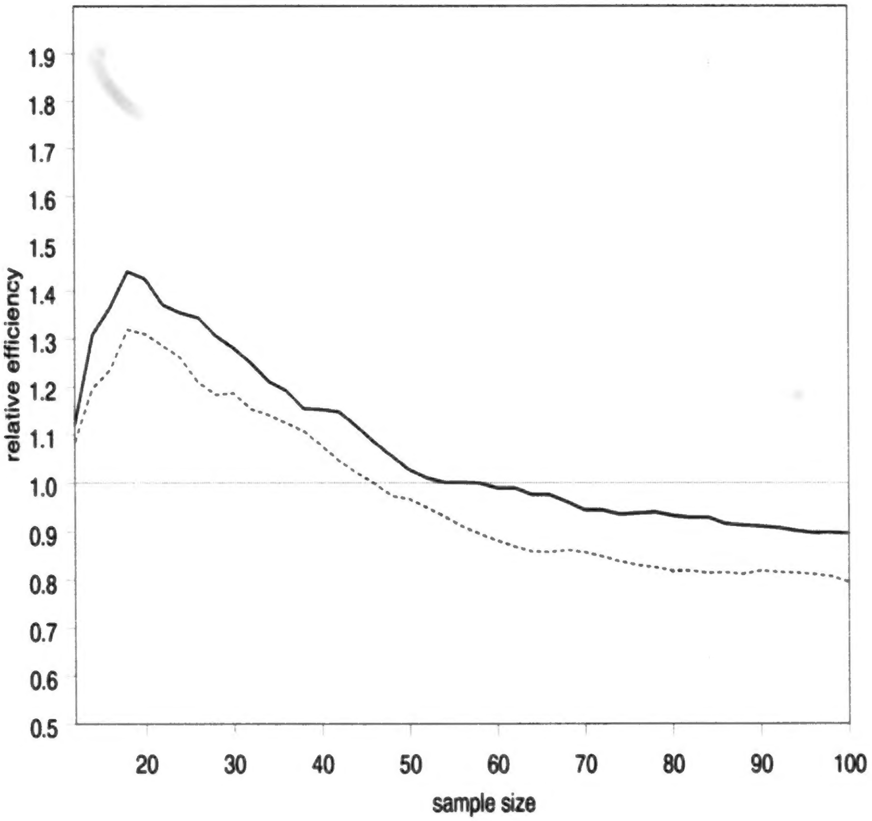

Figure 3.10: Relative efficiency curves for adaptive two-treatment three-period designs when $M = 3$ and $\rho_1 = \rho_3 = \sqrt{\rho_2} = .3$ (the solid line), and $\rho_1 = \rho_3 = \sqrt{\rho_2} = .7$ (the dashed line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$).

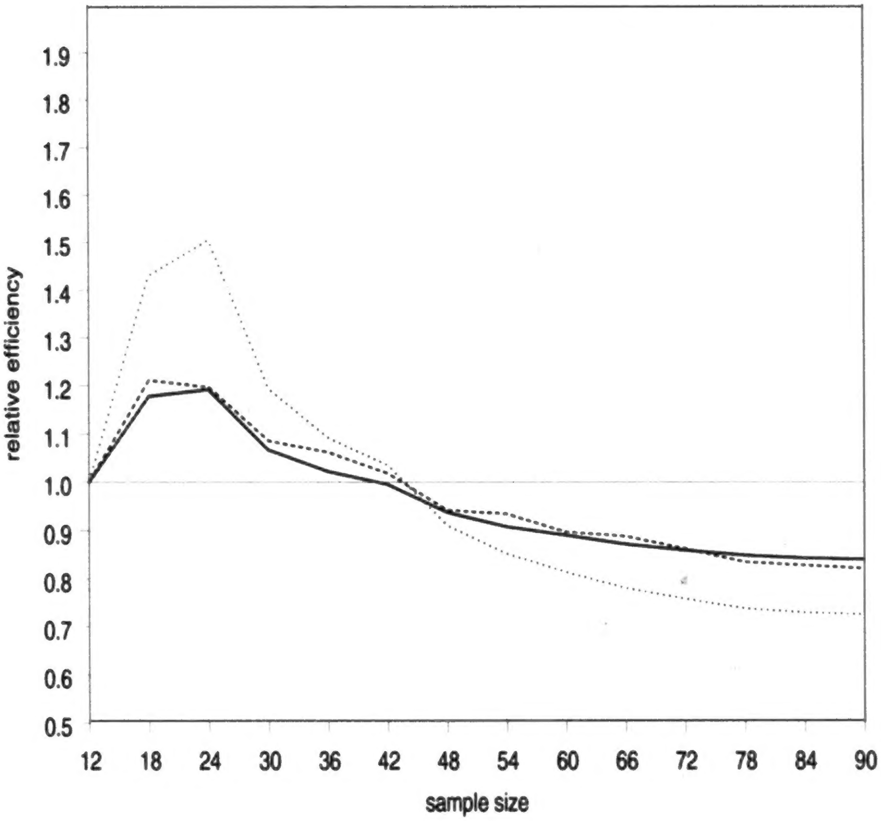

Figure 3.11: Relative efficiency curves for adaptive three-treatment three-period designs when $M = 2$ and correlation coefficients $\rho_1 = \rho_3 = \sqrt{\rho_2} = .3$ and $c_1 = c_2 = 2$ using A-optimal criteria (the solid line), E-optimal criteria (the dashed line) and D-optimal criteria (the dotted line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$).

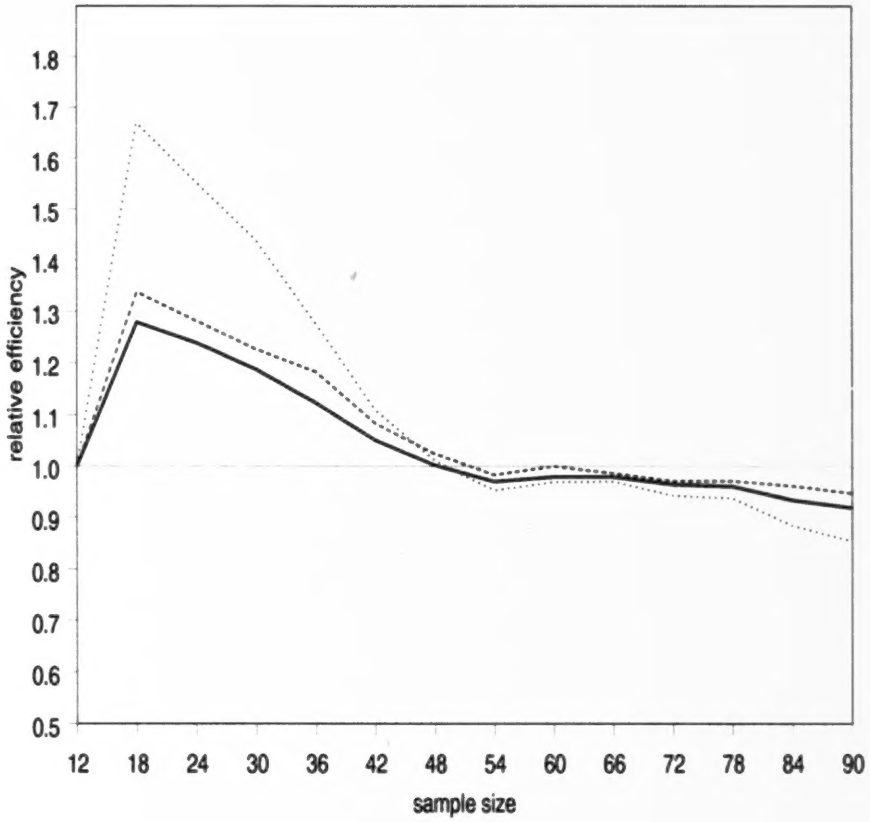

Figure 3.12: Relative efficiency curves for adaptive three-treatment three-period designs when $M = 2$ and correlation coefficients $\rho_1 = \rho_3 = \sqrt{\rho_2} = .3$ and $c_1 = 1, c_2 = 2$ using A-optimal criteria (the solid line), E-optimal criteria (the dashed line) and D-optimal criteria (the dotted line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$).

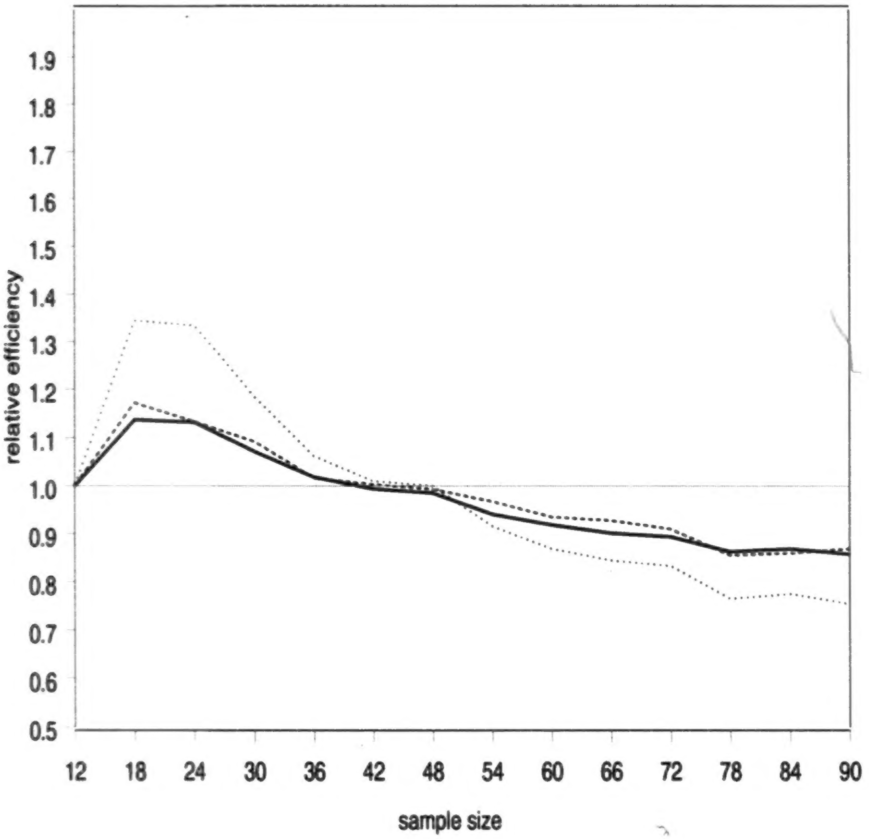

Figure 3.13: Relative efficiency curves for adaptive three-treatment three-period designs when $M = 2$ and correlation coefficients $\rho_1 = \rho_2 = \rho_3 = .3$ and $c_1 = c_2 = 2$ using A-optimal criteria (the solid line), E-optimal criteria (the dashed line) and D-optimal criteria (the dotted line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$).

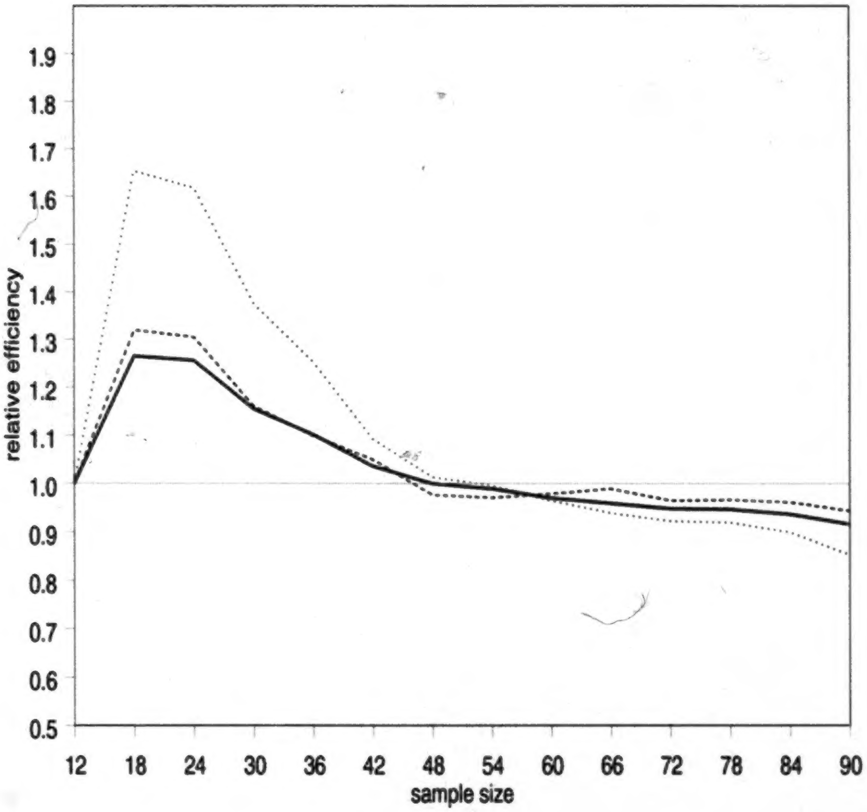

Figure 3.14: Relative efficiency curves for adaptive three-treatment three-period designs when $M = 2$ and correlation coefficients $\rho_1 = \rho_2 = \rho_3 = .3$ and $c_1 = 1, c_2 = 2$ using A-optimal criteria (the solid line), E-optimal criteria (the dashed line) and D-optimal criteria (the dotted line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$).

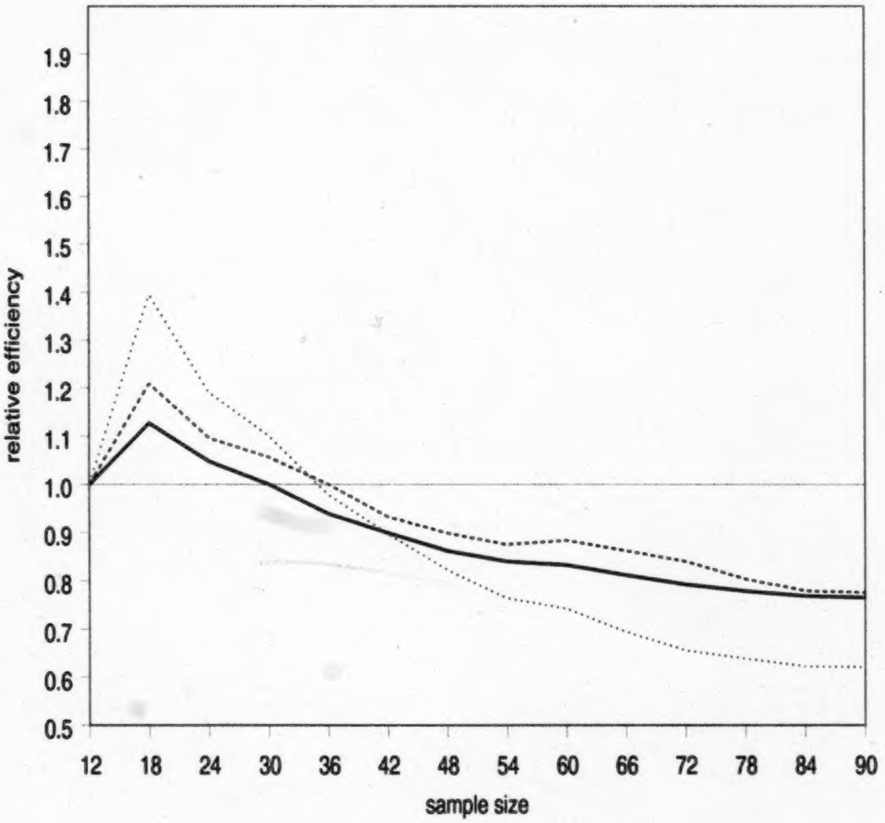

Figure 3.15: Relative efficiency curves for adaptive three-treatment three-period designs when $M = 2$ and correlation coefficients $\rho_1 = \rho_3 = \sqrt{\rho_2} = .7$ and $c_1 = c_2 = 2$ using A-optimal criteria (the solid line), E-optimal criteria (the dashed line) and D-optimal criteria (the dotted line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$).

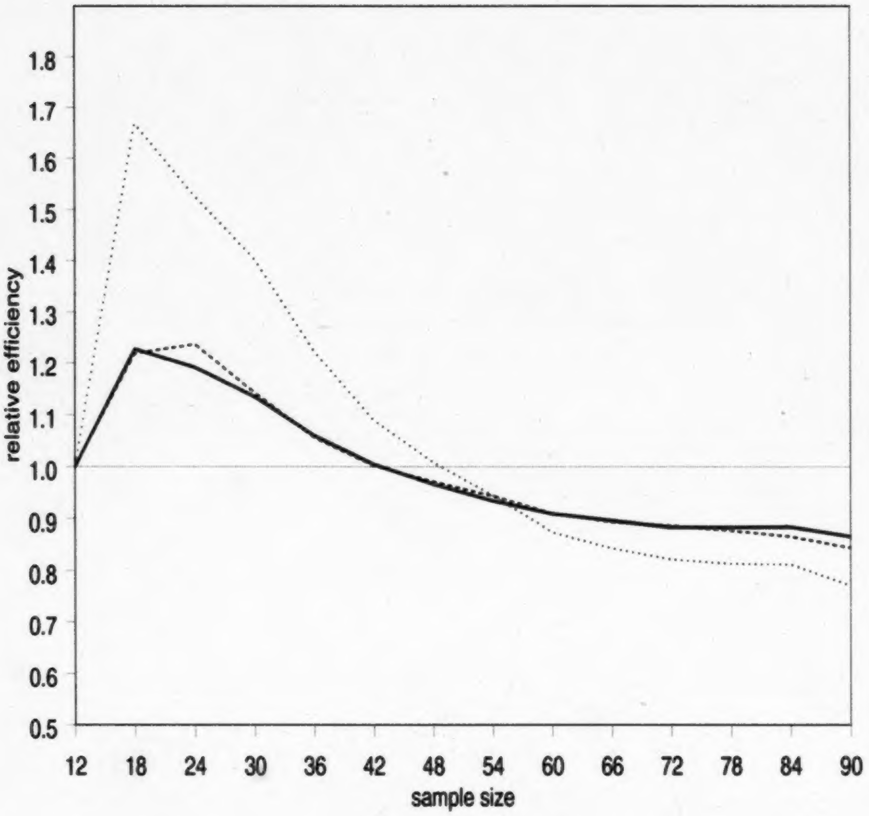

Figure 3.16: Relative efficiency curves for adaptive three-treatment three-period designs when $M = 2$ and correlation coefficients $\rho_1 = \rho_3 = \sqrt{\rho_2} = .7$ and $c_1 = 1, c_2 = 2$ using A-optimal criteria (the solid line), E-optimal criteria (the dashed line) and D-optimal criteria (the dotted line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$).

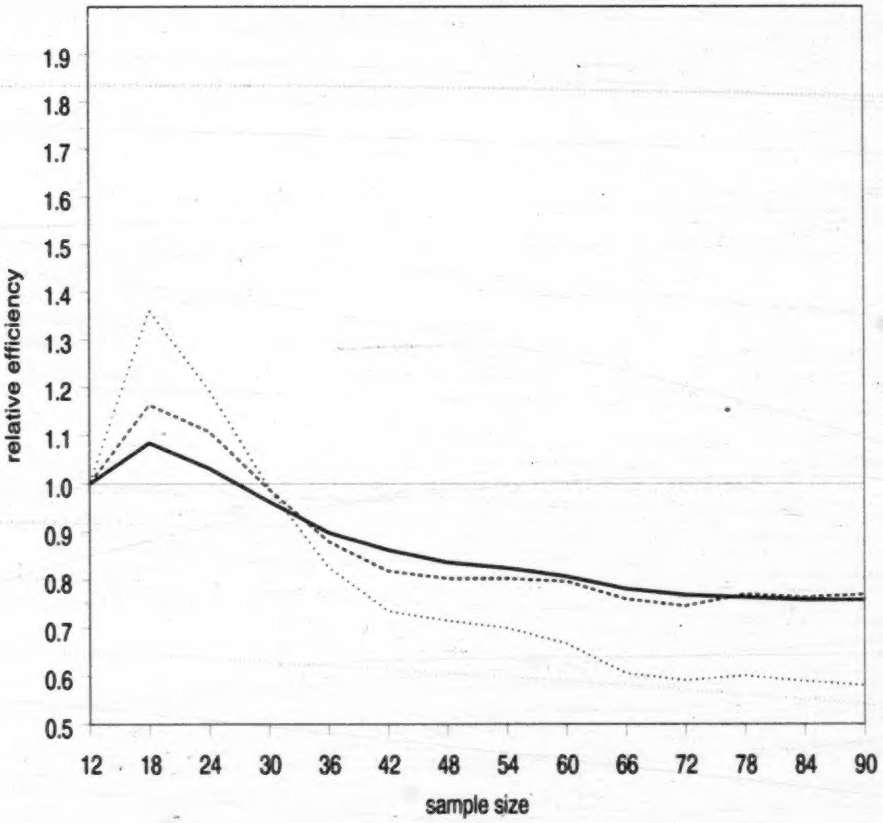

Figure 3.17: Relative efficiency curves for adaptive three-treatment three-period designs when $M = 2$ and correlation coefficients $\rho_1 = \rho_2 = \rho_3 = .7$ and $c_1 = c_2 = 2$ using A-optimal criteria (the solid line), E-optimal criteria (the dashed line) and D-optimal criteria (the dotted line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$).

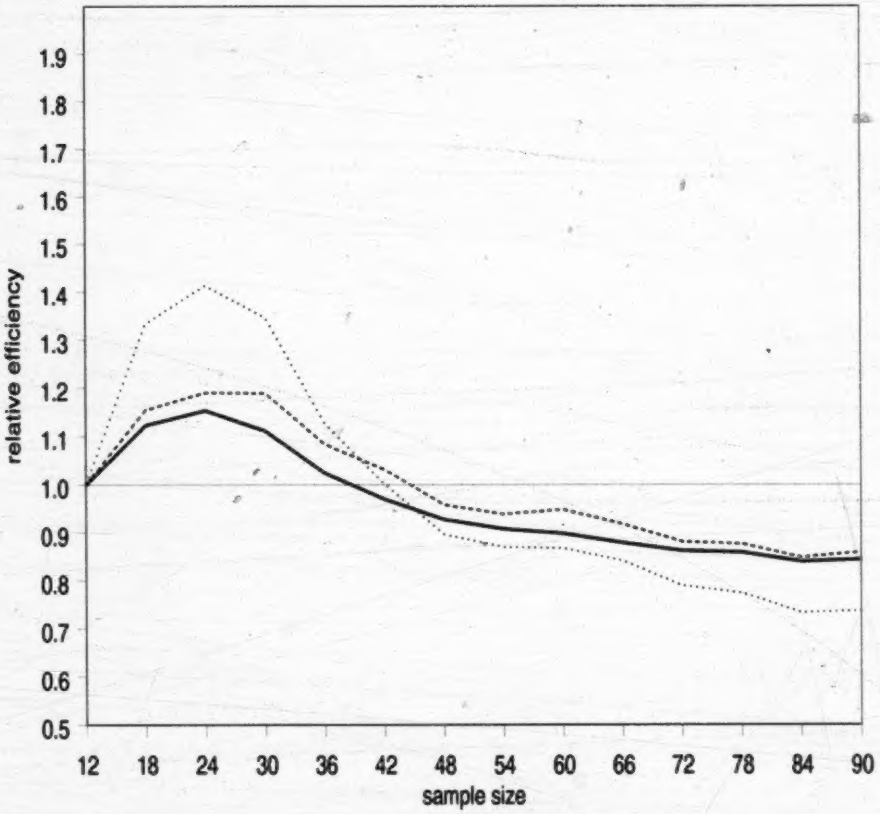

Figure 3.18: Relative efficiency curves for adaptive three-treatment three-period designs when $M = 2$ and correlation coefficients $\rho_1 = \rho_2 = \rho_3 = .7$ and $c_1 = 1, c_2 = 2$ using A-optimal criteria (the solid line), E-optimal criteria (the dashed line) and D-optimal criteria (the dotted line). The straight horizontal line is inserted to indicate the comparison of efficiency improvement (i.e., below the line) relative to the homogeneity case ($M = 1$).

Chapter 4

Use of Proxy Data: An Alternative to Incomplete Data Analysis

4.1 Introduction

Responses from repeated measurement designs make up longitudinal or repeated measures data. In many scientific investigations on repeated measures data, complete data sets are unavailable for various reasons. In this chapter, we investigate analytic methods for repeated measures data with missing values.

Three approaches have been used for dealing with missing data: (1) use only the complete subset data, (2) use an incomplete data analysis method or

(3) use an imputation strategy.

Using only complete subset data has been widely shown to be an inefficient method and to produce inconsistent estimators when missing data are not missing completely at random (Schafer, 1997). Many investigators have developed efficient incomplete data analysis methods by using all available data (for example, Carriere, 1994a and 1999). Alternatively, pseudo values can be imputed in place of the missing data in order to apply a standard complete data analysis method. Treating these values as if they were actual, a repeated-imputation inference can be used to draw conclusions (Rubin, 1987).

In this chapter, we propose yet another strategy for dealing with incomplete data. Analysts often collect proxy information in an attempt to create a complete data situation with no missing values. For example, in cancer clinical trials, some patients may be too sick to respond, but the researcher often takes approximate information from their care providers. Then, the issue is how to treat this proxy information.

Various authors (Grootendorst et al., 1997; Jalukar et al., 1998) have shown that there is a high degree of agreement in the responses of the study subjects and their proxies (care providers). Most of the related theoretical work to date, however, is limited to situations where incomplete information occurs on covariates, and there has been much debate about the optimality of the approaches that omit unobservable covariates instead of utilizing their proxies.

McCallum (1972) and Wickens (1972) demonstrated that, in terms of asymptotic bias for the case of one unobservable regressor, using even a poor proxy is better than using only the observable regressors. Barnow (1976) showed that when there is more than one unobservable regressor, deleting the regressors may be a better choice. Aigner (1974) found that proxies are preferable in most empirical situations in terms of scalar-valued mean squared error (MSE). However, Maddala (1977) showed that including the proxy variable may result in non-negligible bias. Frost (1979) found that using proxy information indiscriminately may be risky from the viewpoint of MSE. Dhrymes (1978), Ohtani (1981), and Trenkler and Stahlecker (1996) outlined situations where estimators with proxy data are inferior to those without, with respect to MSE-matrix criterion.

Recently, investigations on the use of proxy data have been carried on for the experiments using double sampling, where incomplete observations on the variable of interest (either dependent or independent variable) are supplemented by fully observed proxy data. Engel and Walstra (1991) improved the precision for estimating the parameters of interest by a double regression: one with the response and proxy variable on covariates and the other with the response variable on the proxy data. The method they proposed was restricted to a univariate case for large sample studies. Pepe (1992) and Pepe et al. (1994) proposed analysis methods for discrete response variable by solving some estimating equations. Later, Reilly and Pepe (1995) extended the methods to the problem of auxiliary covariate data.

The foremost advantage of incorporating proxy information is that it avoids the missing data problem so that standard statistical analysis methods can be applied. Further, this approach can lead to substantial cost savings when the procedure for obtaining actual data is expensive, if not impossible. In the situation we consider, the dependent data are available, but some data are proxy, not actual. Here, we do not require that the proxy data are available for all study subjects, as the methods developed in the setting of double sampling. We assume that the missing data occur at random (Rubin, 1976). We investigate the merits of using proxy data in place of missing values in terms of efficiency and power in a multivariate repeated measures data setting.

The organization of this chapter is as follows. Section 2 presents the model for incorporating proxy information. Section 3 presents a brief review of incomplete data analysis methods. Section 4 estimates the parameters of interest under various models discussed in Section 2. In Section 5, we propose testing procedures for the method of using proxy. Then, in Section 6, we compare the power of the estimators using proxy data to those without. Finally, in Section 7, we specify the conditions under which the estimators utilizing proxy information are nearly as efficient as those using the complete actual data for selected two and three-period two-treatment repeated measurement designs.

4.2 Model

We consider repeated measurement designs where data are measured over p periods from each subject in s groups (treatment sequences). Let y_{ijk} and p_{ijk} denote the data and their proxy obtained in period i from subject j in sequence k, $i = 1, \ldots, p$, $j = 1, \ldots, N_k$, $k = 1, \ldots, s$. The design structure is given by the matrix $\mathbf{X}_k$, the same for all subjects in sequence k as in Chapter 2.

Let $\mathbf{z} = (\mathbf{z}_1^T, \ldots, \mathbf{z}_s^T)^T$ with $\mathbf{z}_k = (\mathbf{z}_{1k}^T, \ldots, \mathbf{z}_{N_k k}^T)^T$ and $\mathbf{z}_{jk} = (z_{1jk}, \ldots, z_{pjk})^T$, where $z_{ijk} = y_{ijk}$ if y_{ijk} is observed from the study subject, $z_{ijk} = p_{ijk}$ if y_{ijk} is observed from its proxy. Define $\mathbf{M}_{jk} = diag(\delta_{1jk}, \ldots, \delta_{pjk})$ with $\delta_{ijk} = 1$ if y_{ijk} is missing, and 0, otherwise, and $\mathbf{M} = Diag(\mathbf{M}_{11}, \ldots, \mathbf{M}_{N_s s})$. Then $\delta_{.jk} = \sum_{i=1}^{p} \delta_{ijk}$ is the number of periods with proxy information for subject j in sequence k. Let $\mathbf{D}_1 = Diag(\mathbf{1}_{[N_1]} \otimes \mathbf{I}_{[p]}, \ldots, \mathbf{1}_{[N_s]} \otimes \mathbf{I}_{[p]})$ so that $\mathbf{D} = (\mathbf{D}_1, \mathbf{M}\mathbf{D}_1)$ is the design matrix under the cell means model. Then the model we consider is

$$\mathbf{z} = \mathbf{D}\boldsymbol{\omega} + \boldsymbol{\varepsilon} + \mathbf{M}\mathbf{e} = \begin{pmatrix} \mathbf{D}_1 & \mathbf{M}\mathbf{D}_1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\nu} \end{pmatrix} + \boldsymbol{\varepsilon} + \mathbf{M}\mathbf{e} \qquad (4.2.1)$$

The matrix $\mathbf{X} = (\mathbf{X}_1^T, \ldots, \mathbf{X}_s^T)^T$ relates to $\boldsymbol{\beta} = (\mu,\ \boldsymbol{\pi}^T,\ \boldsymbol{\tau}^T,\ \boldsymbol{\gamma}^T,\ \boldsymbol{\lambda}^T)^T$, where μ is the overall mean, $\boldsymbol{\pi}$ the period effects, $\boldsymbol{\tau}$ the treatment effects, $\boldsymbol{\gamma}$ the residual treatment effects and $\boldsymbol{\lambda}$ the sequence effects, with $\mathbf{X}_k = (\mathbf{x}_{1k}, \ldots, \mathbf{x}_{pk})^T$ for $\mathbf{x}_{ik} = (x_{ik1}, \ldots, x_{ikq})^T$ with q the length of $\boldsymbol{\beta}$.

Let $\boldsymbol{\eta} = (\mu_a,\ \boldsymbol{\pi}_a{}^T,\ \boldsymbol{\tau}_a{}^T,\ \boldsymbol{\gamma}_a{}^T,\ \boldsymbol{\lambda}_a{}^T)^T$ measure possible bias from $\boldsymbol{\beta}$ due to proxy. Then $\boldsymbol{\mu} = \mathbf{X}\ \boldsymbol{\beta}$, $\boldsymbol{\nu} = \mathbf{X}\ \boldsymbol{\eta}$ and $\boldsymbol{\omega} = (\boldsymbol{\mu}^T, \boldsymbol{\nu}^T)^T$. The $\boldsymbol{\varepsilon}$ has a mean of $\mathbf{0}$ with covariance matrix $\mathbf{V}$, where $\mathbf{V} = \mathbf{I}_{[N]} \otimes \boldsymbol{\Sigma}$ for a $p \times p$ covariance matrix $\boldsymbol{\Sigma}$ associated with completely observed subjects, with $N = \sum_k N_k$. The extra error term $\mathbf{e}$ for proxy data has a mean of $\mathbf{0}$ with covariance matrix $\mathbf{V_e}$, where $\mathbf{V}_e = \mathbf{I}_{[N]} \otimes \boldsymbol{\Sigma}_e$ for a $p \times p$ covariance matrix $\boldsymbol{\Sigma}_e$ associated with proxy data, and $Cov(\boldsymbol{\varepsilon}, \mathbf{e}) = \mathbf{0}$. Thus, we have $Cov(\mathbf{z}_{jk}) = \boldsymbol{\Sigma} + \mathbf{M}_{jk}\boldsymbol{\Sigma}_e\mathbf{M}_{jk}$. Note here that, when $\mathbf{e} = \mathbf{0}$, the subjects with complete observations and those with proxy data share the same covariance matrix.

The parameters of interest are $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$. When $\mathbf{X}$ is invertible, $\hat{\boldsymbol{\beta}} = \mathbf{X}^{-1}\hat{\boldsymbol{\mu}}$ and $Cov(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T Cov(\hat{\boldsymbol{\mu}})^{-1}\mathbf{X})^{-1} = \mathbf{X}^{-1} Cov(\hat{\boldsymbol{\mu}})(\mathbf{X}^T)^{-1}$, similar to the work of Carriere (1994a) with no proxy data.

Note that the covariance matrix of $\hat{\boldsymbol{\beta}}$ depends on the assumed error structure and the number and the pattern of proxy observations. The proxy pattern is given by matrices $\mathbf{M}_{jk}$, $j = 1, \ldots, N_k$, and $k = 1, \ldots, s$. We assume that at least the first period is completely observed and the missing pattern is monotonic. Each of $\mathbf{M}_{jk}$ might take values $Diag(\mathbf{0}_{[R_l \times R_l]}, \mathbf{I}_{[p-R_l]})$ for $l = 1, \ldots, L$ for missing data occurring at L levels with $0 \leq R_1 < R_2 \ldots < R_L < p$. Thus, $N_k^{(l)} = N_k - \delta_{R_l.k}$ subjects complete the first R_l periods. Note $N_k^{(1)} = N_k$ and

$N^{(l)} = \sum_k N_k^{(l)}$. Denote $p_l = R_l - R_{l-1}$ as the number of periods in the l^{th} level. We order $\{\mathbf{z}_{jk}\}$ ascending with $\delta_{.jk}$, so that the first $N_k^{(l)}$ subjects complete the first R_l periods.

For complete data with no proxy observations, $\boldsymbol{\eta} = \mathbf{0}$, $\mathbf{e} = \mathbf{0}$ and $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\beta}}$ reduces to the usual MLE $\hat{\boldsymbol{\beta}}^f = (\sum_k N_k \mathbf{X}_k^T \mathbf{\Sigma}^{-1} \mathbf{X}_k)^{-1} \sum_k \mathbf{X}_k^T \ \mathbf{\Sigma}^{-1} \mathbf{y}_{.k}$, with $\mathbf{y}_{.k} = \sum_j \mathbf{y}_{jk}$, and its covariance matrix $Cov(\hat{\boldsymbol{\beta}}^f) = (\sum_k N_k \mathbf{X}_k^T \mathbf{\Sigma}^{-1} \mathbf{X}_k)^{-1}$, as in Carriere (1994a), where the superscript "f" denotes that the estimator is based on the "fully actual data." For complete subset data analysis denoted with the superscript "c", the estimator $\hat{\boldsymbol{\beta}}^c$ and its covariance matrix $cov(\hat{\boldsymbol{\beta}}^c)$ are obtained by removing the incomplete pairs of data and are similar in form to $\hat{\boldsymbol{\beta}}^f$ and $Cov(\hat{\boldsymbol{\beta}}^f)$.

4.3 Incomplete data analysis

This section reviews and establishes the relationship of our approach to that of Carriere (1994a and 1999). Carriere (1994a and 1999) considered the model (4.2.1) with $\mathbf{e} = \mathbf{0}$ and no proxy situation. Let $\overline{y}_{i.k}^{(l)} = \sum_{j=1}^{N_k^{(l)}} y_{ijk}/N_k^{(l)}$ for $i = R_{l-1} + 1, \ldots, R_l$, $\overline{\mathbf{y}}_{(p_l).k}^{(l)} = (\overline{y}_{R_{l-1}+1.k}^{(l)}, \ldots, \overline{y}_{R_l.k}^{(l)})^T$, $l = 1, \ldots, L$, $R_0 = 0$, and $\overline{\mathbf{y}}_{k(R_{l-1})}^{(l)} = (\overline{\mathbf{y}}_{(p_1).k}^{(l)\,T}, \ldots, \overline{\mathbf{y}}_{(p_{l-1}).k}^{(l)\,T})^T$. Denote the means $\boldsymbol{\mu}_{k(R_{l-1})} = (\boldsymbol{\mu}_{(p_1)k}^T, \ldots, \boldsymbol{\mu}_{(p_{l-1}),k}^T)^T$, where $\boldsymbol{\mu}_{(p_l)k} = (\mu_{R_{l-1}+1,k}, \ldots, \mu_{R_l,k})^T$, $l = 1, \ldots, L$. Partition $\mathbf{\Sigma}$ to $\mathbf{\Sigma}_{11}^{(l)}$, $\mathbf{\Sigma}_{12}^{(l)}$ and $\mathbf{\Sigma}_{22}^{(l)}$ to divide the subjects between those with complete data up to $(l-1)^{th}$ periods and those with incomplete data at the l^{th} period. Result (4.3.1) states the MLE that Carriere (1999) obtained for the means of $\mathbf{y}$ in each period and sequence with missing data.

Result 4.3.1 *(Carriere, 1999)*

When $\boldsymbol{\Sigma}$ *is known, the MLE for the means* $\boldsymbol{\mu}_k = (\boldsymbol{\mu}^T_{(p_1)k}, \ldots, \boldsymbol{\mu}^T_{(p_L)k})^T$, $k = 1, \ldots, s$ *is*

$$\hat{\boldsymbol{\mu}}_{(p_l)k} = \overline{\mathbf{y}}^{(l)}_{(p_l).k} - \boldsymbol{\Sigma}^{(l)}_{21}(\boldsymbol{\Sigma}^{(l)}_{11})^{-1}(\overline{\mathbf{y}}^{(l)}_{k(R_{l-1})} - \hat{\boldsymbol{\mu}}_{k(R_{l-1})}) \tag{4.3.1}$$

for $l = 1, \ldots, L$*. When* $\boldsymbol{\Sigma}$ *is unknown, the MLE for the means is estimated as*

$$\hat{\boldsymbol{\mu}}_{(p_l)k} = \overline{\mathbf{y}}^{(l)}_{(p_l).k} - \mathbf{S}^{(l)}_{21}(\mathbf{S}^{(l)}_{11})^{-1}(\overline{\mathbf{y}}^{(l)}_{k(R_{l-1})} - \hat{\boldsymbol{\mu}}_{k(R_{l-1})}) \tag{4.3.2}$$

where $\mathbf{S}$ *is the sample covariance matrix corresponding to* $\boldsymbol{\Sigma}$*. Partitioning* $\mathbf{S}$ *in the same way as partitioning* $\boldsymbol{\Sigma}$ *gives* $\mathbf{S}^{(l)}_{11}$, $\mathbf{S}^{(l)}_{12}$ *and* $\mathbf{S}^{(l)}_{22}$.

Remark 4.3.1 *When* $L = 2$*, Carriere (1999) obtained that*

$$\hat{\boldsymbol{\mu}}_{(p_2)k} = \overline{\mathbf{y}}^{(2)}_{(p_2).k} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}^{-1}_{11}(\overline{\mathbf{y}}^{(2)}_{(p_1).k} - \overline{\mathbf{y}}^{(1)}_{(p_1).k}) \tag{4.3.3}$$

and $\hat{\boldsymbol{\mu}}_{(p_1)k} = \overline{\mathbf{y}}^{(1)}_{(p_1).k}$*. The covariance matrix for* $\hat{\boldsymbol{\mu}}_k$ *are* $Cov(\hat{\boldsymbol{\mu}}_{(p_1)k}) = \boldsymbol{\Sigma}_{11}/N_k$, $Cov(\hat{\boldsymbol{\mu}}_{(p_1)k}, \hat{\boldsymbol{\mu}}_{(p_2)k}) = \boldsymbol{\Sigma}_{12}/N_k$ *and*

$$Cov(\hat{\boldsymbol{\mu}}_{(p_2)k}) = \frac{1}{N^{(2)}_k}[\boldsymbol{\Sigma}_{22} - \frac{N_k - N^{(2)}_k}{N_k}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}^{-1}_{11}\boldsymbol{\Sigma}_{12}] \tag{4.3.4}$$

for $\boldsymbol{\Sigma}$ is known. When $\boldsymbol{\Sigma}$ is unknown, the covariance for $\hat{\boldsymbol{\mu}}_{(p_2)k}$ becomes

$$Cov(\hat{\boldsymbol{\mu}}_{(p_2)k}) = \frac{1}{N_k^{(2)}}\{\boldsymbol{\Sigma}_{22} + \frac{N_k - N_k^{(2)}}{N_k}[\frac{p_1}{N^{(2)} - s - p_1 - 1}\boldsymbol{\Sigma}_{2.1} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}]\} \tag{4.3.5}$$

reflecting the sampling variation in estimating $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}_{2.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$.

Remark 4.3.2 *For $p_2 = 1$, Carriere (1999) observed that*

$$\begin{aligned} Cov(\hat{\boldsymbol{\mu}_{2k}}) &= \frac{\sigma_{pp}}{N_k^{(2)}}\{1 + \\ &\quad \frac{N_k - N_k^{(2)}}{N_k}[\frac{p_1}{N(1-r) - s - p_1 - 1}(1 - \frac{\boldsymbol{\delta}_{21}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\delta}_{21}}{\sigma_{pp}}) - \frac{\boldsymbol{\delta}_{21}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\delta}_{21}}{\sigma_{pp}}]\} \end{aligned} \tag{4.3.6}$$

where $\boldsymbol{\delta}_{21} = (\sigma_{p1}, \ldots, \sigma_{pp_1})^T$ and r is the proportion of missing data.

4.4 Parameter estimation with proxy data

4.4.1 When $\mathbf{e} = \mathbf{0}$

When $\mathbf{e} = \mathbf{0}$, the proxy is assumed to behave similarly to the actual data, but possibly at a different location (i.e., $\boldsymbol{\nu} \neq \mathbf{0}$). Let $\mathbf{y}_{jk(R_{l-1})} = (\mathbf{y}_{(p_1)jk}^T, \ldots, \mathbf{y}_{(p_{l-1})jk}^T)^T$ with $\mathbf{y}_{(p_l)jk} = (y_{R_{l-1}+1,jk}, \ldots, y_{R_l,jk})^T$, $l = 1, \ldots, L$. Similarly, $\mathbf{p}_{jk(R_{l-1})} = (\mathbf{p}_{(p_1)jk}^T, \ldots, \mathbf{p}_{(p_{l-1})jk}^T)^T$ with $\mathbf{p}_{(p_l)jk} = (p_{R_{l-1}+1,jk}, \ldots, p_{R_l,jk})^T$, $l = 2, \ldots, L$.

The likelihood can be written as

$$\mathcal{L} = L_1 L_2 \dots L_L \tag{4.4.1}$$

where

$$L_1 = \prod_{jk} f(\mathbf{y}_{(p_1)jk}), \tag{4.4.2}$$

$$L_2 = \prod_{k} \prod_{j=1}^{N_k^{(2)}} f(\mathbf{y}_{(p_2)jk} | \mathbf{y}_{(p_1)jk}) \prod_{k} \prod_{j=N_k^{(2)}+1}^{N_k} f(\mathbf{p}_{(p_2)jk} | \mathbf{y}_{(p_1)jk}) \tag{4.4.3}$$

and more generally, for $l = 1, \dots, L$,

$$L_l = \prod_{k} \prod_{j=1}^{N_k^{(l)}} f(\mathbf{y}_{(p_l)jk} | \mathbf{y}_{jk(R_{l-1})}, \mathbf{p}_{jk(R_{l-1})}) \prod_{k} \prod_{j=N_k^{(l)}+1}^{N_k} f(\mathbf{p}_{(p_l)jk} | \mathbf{y}_{jk(R_{l-1})}, \mathbf{p}_{jk(R_{l-1})}) \tag{4.4.4}$$

Successively maximizing $L_1, \dots, L_L$ gives the MLEs for $\boldsymbol{\mu}_{(p_l)k}$, $l = 1, \dots, L$ and $\boldsymbol{\nu}_{(p_l)k}$, $l = 2, \dots, L$ when $\mathbf{\Sigma}$ is known as

$$\hat{\boldsymbol{\mu}}_{(p_1)k} = \overline{\mathbf{y}}^{(1)}_{(p_1).k} \tag{4.4.5}$$

$$\hat{\boldsymbol{\mu}}_{(p_l)k} = \overline{\mathbf{y}}^{(l)}_{(p_l).k} - \mathbf{B}^{(l)} (\overline{\mathbf{y}}^{(l)}_{k(R_{l-1})} - \hat{\boldsymbol{\mu}}_{k(R_{l-1})}) \tag{4.4.6}$$

and

$$\hat{\boldsymbol{\nu}}_{(p_l)k} = \overline{\mathbf{p}}^{(l)}_{(p_l).k} - \overline{\mathbf{y}}^{(l)}_{(p_l).k} + \frac{N_k}{N_k - N_k^{(l)}} \mathbf{B}^{(l)} (\overline{\mathbf{y}}^{(l)}_{k(R_{l-1})} - \hat{\boldsymbol{\mu}}_{k(R_{l-1})}), \tag{4.4.7}$$

where $\overline{p}_{i.k}^{(l)} = \sum_{j=N_k^{(l)}+1}^{N_k} p_{ijk}/(N_k - N_k^{(l)})$ for $i = R_{l-1}+1, \ldots, R_l$, $\overline{\mathbf{p}}_{(p_l).k}^{(l)} = (\overline{p}_{R_{l-1}+1.k}^{(l)}, \ldots, \overline{p}_{R_l.k}^{(l)})^T$, $l = 2, \ldots, L$, and $\overline{\mathbf{p}}_{k(R_{l-1})}^{(l)} = (\overline{\mathbf{p}}_{(p_1).k}^{(l)}{}^T, \ldots, \overline{\mathbf{p}}_{(p_{l-1}).k}^{(l)}{}^T)^T$ for $k = 1, \ldots, s$ and $\mathbf{B}^{(l)} = \mathbf{\Sigma}_{21}^{(l)} \mathbf{\Sigma}_{11}^{(l)^{-1}}$.

When $\mathbf{\Sigma}$ is unknown, the MLE for $\mathbf{B}^{(l)}$ is

$$
\begin{aligned}
\hat{\mathbf{B}}^{(l)} = & \sum_{k=1}^{s}\Big[\sum_{j=1}^{N_k^{(l)}}(\mathbf{y}_{(p_l)jk} - \overline{\mathbf{y}}_{(p_l).k}^{(l)})(\mathbf{y}_{jk(R_{l-1})} - \overline{\mathbf{y}}_{k(R_{l-1})}^{(l)})^T \\
& + \sum_{i=0}^{l-2} \sum_{j=N_k^{(l-i)}+1}^{N_k^{(l-i-1)}} (\mathbf{p}_{(p_l)jk} - \overline{\mathbf{p}}_{(p_l).k}^{(l)}) \\
& \times \ (\tilde{\mathbf{y}}_{jk(R_{l-1})}^{(i)} + \frac{N_k^{(l)}}{N_k - N_k^{(l)}}(\overline{\mathbf{y}}_{k(R_{l-1})}^{(l)} - \hat{\boldsymbol{\mu}}_{k(R_{l-1})}))^T\Big] \\
& \times \ \Big[\sum_{k=1}^{s}\Big\{\sum_{j=1}^{N_k^{(l-1)}} (\mathbf{y}_{jk(R_{l-1})} - \overline{\mathbf{y}}_{k(R_{l-1})}^{(l)})(\mathbf{y}_{jk(R_{l-1})} - \overline{\mathbf{y}}_{k(R_{l-1})}^{(l)})^T \\
& + \sum_{i=0}^{l-2} \sum_{j=N_k^{(l-i)}+1}^{N_k^{(l-i-1)}} (\tilde{\mathbf{y}}_{jk(R_{l-1})}^{(i)} + \frac{N_k^{(l)}}{N_k - N_k^{(l)}}(\overline{\mathbf{y}}_{k(R_{l-1})}^{(l)} - \hat{\boldsymbol{\mu}}_{k(R_{l-1})})) \\
& \times \ (\tilde{\mathbf{y}}_{jk(R_{l-1})}^{(i)} + \frac{N_k^{(l)}}{N_k - N_k^{(l)}}(\overline{\mathbf{y}}_{k(R_{l-1})}^{(l)} - \hat{\boldsymbol{\mu}}_{k(R_{l-1})}))^T\Big\}\Big]^{-1}, \qquad (4.4.8)
\end{aligned}
$$

where $\tilde{\mathbf{y}}_{jk(R_{l-1})}^{(i)}$ is the vector $\mathbf{y}_{jk(R_{l-1})} - \hat{\boldsymbol{\mu}}_{k(R_{l-1})}$ with the last i components replaced by $((\mathbf{p}_{(p_{l-i})jk} - \hat{\boldsymbol{\mu}}_{(p_{l-i})k} - \hat{\boldsymbol{\nu}}_{(p_{l-i})k})^T, \ldots, (\mathbf{p}_{(p_{l-1})jk} - \hat{\boldsymbol{\mu}}_{(p_{l-1})k} - \hat{\boldsymbol{\nu}}_{(p_{l-1})k})^T)^T$ and $\tilde{\mathbf{y}}_{jk(R_{l-1})}^{(0)} = \mathbf{y}_{jk(R_{l-1})} - \hat{\boldsymbol{\mu}}_{k(R_{l-1})}$. The covariance matrices for the means (4.4.5), (4.4.6) and (4.4.7) can be obtained in a straightforward manner.

Remark 4.4.1 *When $\boldsymbol{\nu} = \mathbf{0}$, the data are complete with no missing or proxy values. Thus, the MLE $\hat{\boldsymbol{\mu}}_k$ becomes the usual sample mean (Carriere, 1999).*

Theorem 4.4.1 *When $L = 2$ and $\mathbf{\Sigma}$ is known, we obtain the MLE and covariance matrices for $\boldsymbol{\mu}_{(p_1)k}$, $\boldsymbol{\mu}_{(p_2)k}$ as in Remark 4.3.1. The additional proxy information contributes to the estimation of $\boldsymbol{\nu}_{(p_2)k}$, which is*

$$\hat{\boldsymbol{\nu}}_{(p_2)k} = \overline{\mathbf{p}}^{(2)}_{(p_2).k} - \overline{\mathbf{y}}^{(2)}_{(p_2).k} + \frac{N_k}{N_k - N_k^{(2)}} \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1} (\overline{\mathbf{y}}^{(2)}_{(p_1).k} - \overline{\mathbf{y}}^{(1)}_{(p_1).k}) \qquad (4.4.9)$$

The covariance matrices are

$$Cov(\hat{\boldsymbol{\nu}}_{(p_2)k}) = \frac{N_k}{(N_k - N_k^{(2)}) N_k^{(2)}} \mathbf{\Sigma}_{2.1} \qquad (4.4.10)$$

$$Cov(\hat{\boldsymbol{\nu}}_{(p_2)k}, \hat{\boldsymbol{\mu}}_{(p_1)k}) = \mathbf{0} \qquad (4.4.11)$$

$$Cov(\boldsymbol{\mu}_{(p_2)k}, \boldsymbol{\nu}_{(p_2)k}) = -\frac{\mathbf{\Sigma}_{2.1}}{N_k^{(2)}} \qquad (4.4.12)$$

PROOF. The proof is obtained in a straightforward manner by substituting $l = 2$ in (4.4.5), (4.4.6), (4.4.7) and applying the properties of multivariate normal distribution.

$\square$

Define the sum of squares $\mathbf{SS}^{(l)}_{mn} = \sum_k \sum_{j=1}^{N_k^{(l)}} (\mathbf{y}_{(p_m)jk} - \overline{\mathbf{y}}^{(l)}_{(p_m).k})(\mathbf{y}_{(p_n)jk} - \overline{\mathbf{y}}^{(l)}_{(p_n).k})^T$, $l = 1, 2$, $m = 1, 2$, $n = 1, 2$, $\mathbf{SS}^{(0)}_{11} = \sum_k \sum_{j=N_k^{(2)}+1}^{N_k} (\mathbf{y}_{(p_1)jk} - \overline{\mathbf{y}}^{(0)}_{(p_1).k}) (\mathbf{y}_{(p_1)jk} - \overline{\mathbf{y}}^{(0)}_{(p_1).k})^T$, $\mathbf{SS}^{(0)}_{12} = \sum_k \sum_{j=N_k^{(2)}+1}^{N_k} (\mathbf{y}_{(p_1)jk} - \overline{\mathbf{y}}^{(0)}_{(p_1).k})(\mathbf{p}_{(p_2)jk} - \overline{\mathbf{p}}^{(2)}_{(p_2).k})^T$, $\mathbf{SS}^{(0)}_{22} = \sum_k \sum_{j=N_k^{(2)}+1}^{N_k} (\mathbf{p}_{(p_2)jk} - \overline{\mathbf{p}}^{(2)}_{(p_2).k})(\mathbf{p}_{(p_2)jk} - \overline{\mathbf{p}}^{(2)}_{(p_2).k})^T$, where $\overline{\mathbf{y}}^{(0)}_{(p_1).k} =$

$\sum_{j=N_k^{(2)}+1}^{N_k} \mathbf{y}_{(p_1)jk}/(N_k - N_k^{(2)})$. And, let $\mathbf{SS}_{11} = \mathbf{SS}_{11}^{(2)} + \mathbf{SS}_{11}^{(0)}$, $\mathbf{SS}_{21} = \mathbf{SS}_{21}^{(2)} + \mathbf{SS}_{21}^{(0)}$, $\mathbf{SS}_{22} = \mathbf{SS}_{22}^{(2)} + \mathbf{SS}_{22}^{(0)}$ and $\mathbf{SS}_{2.1} = \mathbf{SS}_{22} - \mathbf{SS}_{21}\mathbf{SS}_{11}^{-1}\mathbf{SS}_{12}$. Then, we have the following theorem.

Theorem 4.4.2 *When $\boldsymbol{\Sigma}$ is unknown for $L = 2$, the MLE for $\boldsymbol{\mu}_{(p_1)k}$ is unchanged. The MLEs for $\boldsymbol{\mu}_{(p_2)k}$ and $\boldsymbol{\nu}_{(p_2)k}$ are obtained by substituting $\hat{\mathbf{B}}^{(2)} = \mathbf{SS}_{21}\mathbf{SS}_{11}^{-1}$ obtained from (4.4.8) with $l = 2$ in (4.4.6) and (4.4.7). The covariance matrices for $\hat{\boldsymbol{\mu}}_{(p_2)k}$ and $\hat{\boldsymbol{\nu}}_{(p_2)k}$ are*

$$Cov(\hat{\boldsymbol{\mu}}_{(p_2)k}) = \frac{\boldsymbol{\Sigma}_{22}}{N_k^{(2)}} + \frac{N_k - N_k^{(2)}}{N_k^{(2)} N_k}\left(\frac{p_1}{N - 2s - p_1 - 1}\boldsymbol{\Sigma}_{2.1} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right) \quad (4.4.13)$$

$$Cov(\hat{\boldsymbol{\nu}}_{(p_2)k}) = \frac{N_k}{(N_k - N_k^{(2)})N_k^{(2)}} \frac{N - 2s - 1}{N - 2s - p_1 - 1}\boldsymbol{\Sigma}_{2.1} \quad (4.4.14)$$

and

$$Cov(\hat{\boldsymbol{\nu}}_{(p_2)k}, \hat{\boldsymbol{\mu}}_{(p_2)k}) = -\frac{N - 2s - 1}{(N - 2s - p_1 - 1)N_k^{(2)}}\boldsymbol{\Sigma}_{2.1}. \quad (4.4.15)$$

The estimates for the covariance matrices can be obtained by plugging in the unbiased estimators for the components of $\boldsymbol{\Sigma}$, which are $\hat{\boldsymbol{\Sigma}}_{11} = \mathbf{SS}_{11}^{(1)}/(N - s)$, $\hat{\boldsymbol{\Sigma}}_{21} = \hat{\mathbf{B}}^{(2)}\hat{\boldsymbol{\Sigma}}_{11}$, $\hat{\boldsymbol{\Sigma}}_{2.1} = \mathbf{SS}_{2.1}/(N - 2s - p_1)$ and $\hat{\boldsymbol{\Sigma}}_{22} = c\mathbf{SS}_{2.1} + d\hat{\mathbf{B}}^{(2)}(\hat{\boldsymbol{\Sigma}}_{11})(\hat{\mathbf{B}}^{(2)})^T$, with $c = [(N - 3s - p_1)(N - 2s - p_1 - 1) - sp_1]/[(N - 3s)(N - 2s - p_1 - 1)(N - 2s - p_1)]$ and $d = (N - s)/(N - 3s)$.

PROOF. Denote

$$\mathbf{SS} = \begin{bmatrix} \mathbf{SS}_{11} & \mathbf{SS}_{12} \\ \mathbf{SS}_{21} & \mathbf{SS}_{22} \end{bmatrix}$$

then **SS** has a Wishart distribution with covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

and degrees of freedom $N - 2s$. A derivation similar to that of Carriere (1999) leads to (4.4.13), (4.4.14) and (4.4.15). That $\hat{\boldsymbol{\Sigma}}_{2.1}$ is unbiased follows from the fact that $\mathbf{SS}_{2.1}$ has the Wishart distribution with covariance matrix $\boldsymbol{\Sigma}_{2.1}$ and the degrees of freedom $N - 2s - p_1$ (Mardia, Kent and Bibby, 1979, p.71). Similarly, that $\hat{\boldsymbol{\Sigma}}_{11}$ is unbiased follows from the fact that $\mathbf{SS}_{11}^{(1)}$ has the Wishart distribution with covariance matrix $\boldsymbol{\Sigma}_{11}$ and the degrees of freedom $N - s$. Rewrite $\mathbf{SS}_{11}^{(1)}$, so that

$$\mathbf{SS}_{11}^{(1)} = \mathbf{SS}_{11} + \mathbf{SS}_0,$$

where

$$\mathbf{SS}_0 = \sum_k \frac{N_k^{(2)}(N_k - N_k^{(2)})}{N_k} (\overline{\mathbf{y}}_{(p_1).k}^{(2)} - \overline{\mathbf{y}}_{(p_1).k}^{(0)})(\overline{\mathbf{y}}_{(p_1).k}^{(2)} - \overline{\mathbf{y}}_{(p_1).k}^{(0)})^T,$$

and $\hat{\mathbf{B}}^{(2)}$ and $\mathbf{SS}_0$ are independent. We have that

$$E(\mathbf{SS}_0) = s\boldsymbol{\Sigma}_{11}.$$

And, a similar derivation to Morrison (1970) leads to

$$E(\hat{\mathbf{B}}^{(2)}) = \mathbf{B},$$

$$E[(\hat{\mathbf{B}}^{(2)} - \mathbf{B})\boldsymbol{\Sigma}_{11}(\hat{\mathbf{B}}^{(2)} - \mathbf{B})^T] = \frac{p_1}{N - 2s - p_1 - 1}\boldsymbol{\Sigma}_{2.1}.$$

Thus,

$$\begin{aligned} E(\hat{\boldsymbol{\Sigma}}_{21}) &= E[\frac{\hat{\mathbf{B}}^{(2)}}{N-s}(\mathbf{SS}_{11} + \mathbf{SS}_0)] \\ &= E(\frac{\mathbf{SS}_{21}}{N-s}) + E(\frac{\hat{\mathbf{B}}^{(2)}\mathbf{SS}_0}{N-s}) \\ &= \frac{\boldsymbol{\Sigma}_{21}(N-2s)}{N-s} + \frac{E(\hat{\mathbf{B}}^{(2)})E(\mathbf{SS}_0)}{N-s} \\ &= \frac{\boldsymbol{\Sigma}_{21}(N-2s)}{N-s} + \frac{s\boldsymbol{\Sigma}_{21}}{N-s} \\ &= \boldsymbol{\Sigma}_{21}. \end{aligned}$$

Let $\hat{\boldsymbol{\Sigma}}_{22} = c\mathbf{SS}_{2.1} + d\hat{\mathbf{B}}^{(2)}\hat{\boldsymbol{\Sigma}}_{11}(\hat{\mathbf{B}}^{(2)})^T$ for some constants c and d. Then, solving

$$\begin{aligned} E(\hat{\boldsymbol{\Sigma}}_{22}) &= c(N - 2s - p_1)\boldsymbol{\Sigma}_{2.1} + d[\frac{E(\hat{\mathbf{B}}^{(2)}\mathbf{SS}_{11}(\hat{\mathbf{B}}^{(2)})^T) + E(\hat{\mathbf{B}}^{(2)}\mathbf{SS}_0(\hat{\mathbf{B}}^{(2)})^T)}{N-s}] \\ &= c(N - 2s - p_1)\boldsymbol{\Sigma}_{2.1} + d\frac{E(\mathbf{SS}_{22} - \mathbf{SS}_{2.1})}{N-s} + d\frac{E(\hat{\mathbf{B}}^{(2)}E(\mathbf{SS}_0)(\hat{\mathbf{B}}^{(2)})^T)}{N-s} \end{aligned}$$

$$
\begin{aligned}
&= c(N-2s-p_1)\boldsymbol{\Sigma}_{2.1} + d\frac{(p_1+2s-N)\boldsymbol{\Sigma}_{2.1} + (N-2s)\boldsymbol{\Sigma}_{22}}{N-s} \\
&+ ds\frac{[\frac{p_1}{N-2s-p_1-1}+1]\boldsymbol{\Sigma}_{2.1} - \boldsymbol{\Sigma}_{22}}{N-s} \\
&= \boldsymbol{\Sigma}_{22}
\end{aligned}
$$

for c and d, we obtain

$$c = \frac{(N-3s-p_1)(N-2s-p_1-1) - sp_1}{(N-3s)(N-2s-p_1)(N-2s-p_1-1)}$$

and

$$d = \frac{N-s}{N-3s}.$$

Hence, Theorem 4.4.2 is proved.

□

Remark 4.4.2 *For $p_2 = 1$, $Cov(\hat{\boldsymbol{\mu}_{2k}})$ is the same as (4.3.6) with $N(1-r)-s-p_1-1$ replaced by $N-2s-p_1-1$. It is clear that utilizing proxy information is more efficient for estimating $\boldsymbol{\mu}_{(p_2)k}$, $k = 1, \ldots, s$, when the proportion of missing data $r > s/N$.*

Theorem 4.4.3 *When $\boldsymbol{\Sigma}$ is compound symmetric with $L = 2$, an approximately unbiased estimator for ρ and an unbiased estimator for σ^2 are obtained as*

$$\hat{\rho} = \frac{2(\sum_{i>i'}^{p}(\mathbf{SS})_{ii'})}{(p-1)(\sum_{1}^{p}(\mathbf{SS})_{ii})} \tag{4.4.16}$$

$$\hat{\sigma}^2 = \frac{\sum_1^p (\mathbf{SS})_{ii}}{p(N-2s)}, \tag{4.4.17}$$

where **SS** *is defined in Theorem 4.4.2. The variances for* $\hat{\sigma}^2$ *and* $\hat{\rho}$ *are obtained as*

$$V(\hat{\sigma^2}) \approx 2\sigma^4 \frac{1+\rho^2(p-1)}{p(N-2s)} \tag{4.4.18}$$

and

$$V(\hat{\rho}) \approx \frac{2(1-\rho)^2[1+\rho(p-1)]^2}{p(p-1)(N-2s)}. \tag{4.4.19}$$

Proof. We estimate ρ by dividing the sum of all nondiagonal elements by the sum of all diagonal elements of the sample sum of squares. Here, we use not only the complete subset data (Carriere, 1994a), but also the proxy, which shares the same covariance structure as the observed data. Denote that $SSS^{(m)} = p\sum\sum(\overline{z}_{.jk} - \overline{z}^{(m)}_{..k})^2$, $SSE^{(m)} = \sum\sum\sum(z_{ijk} - \overline{z}^{(m)}_{i.k})^2 - SSS^{(m)}$, where $\overline{z}_{.jk} = \sum_i z_{ijk}/p$, for $m = 0, 2$. The mean and the summation are over the complete subset for $m = 2$, and over the subset with proxy components for $m = 0$. Let $SSS = SSS^{(2)} + SSS^{(0)}$ and $SSE = SSE^{(2)} + SSE^{(0)}$. Under the multivariate normal assumption, it can be shown that SSS and SSE are independently distributed as $\sigma^2(1+\rho(p-1))\chi^2_{N-2s}$ and $\sigma^2(1-\rho)\chi^2_{(p-1)(N-2s)}$. Then the result follows by a similar derivation of Carriere (1994a).

□

Remark 4.4.3 *When* $\mathbf{\Sigma}$ *is compound symmetric with* $L = 2$, *the MLEs for*

$\boldsymbol{\mu}_{(p_2)k}$ *and* $\boldsymbol{\nu}_{(p_2)k}$ *become*

$$\hat{\boldsymbol{\mu}}_{(p_2)k} = \overline{\mathbf{y}}^{(2)}_{(p_2).k} - \frac{\hat{\rho}}{1+\hat{\rho}(p_1-1)}\mathbf{1}_{[p_2]}\mathbf{1}^T_{[p_1]}(\overline{\mathbf{y}}^{(2)}_{(p_1).k} - \overline{\mathbf{y}}^{(1)}_{(p_1).k})$$

and

$$\hat{\boldsymbol{\nu}}_{(p_2)k} = \overline{\mathbf{p}}^{(2)}_{(p_2).k} - \overline{\mathbf{y}}^{(2)}_{(p_2).k} - \frac{\hat{\rho}}{1+\hat{\rho}(p_1-1)}\mathbf{1}_{[p_2]}\mathbf{1}^T_{[p_1]}(\overline{\mathbf{y}}^{(0)}_{(p_1).k} - \overline{\mathbf{y}}^{(2)}_{(p_1).k}),$$

where $\hat{\rho}$ *is an estimator for* ρ *as given in (4.4.16).*

Remark 4.4.4 *When* $\boldsymbol{\Sigma}$ *is compound symmetric with* $L = 2$, *the covariance for* $\hat{\boldsymbol{\mu}}_{(p_2)k}$, $\hat{\boldsymbol{\nu}}_{(p_2)k}$ *are*

$$Cov(\hat{\boldsymbol{\mu}}_{(p_2)k}) = \frac{\sigma^2}{N^{(2)}_k}\{(1-\rho)\mathbf{I}_{[p_2]}+[\rho+\frac{N_k - N^{(2)}_k}{N_k}p_1\{a_1[1+(p_1-1)\rho]-2a_2\rho\}]\mathbf{1}_{[p_2]}\mathbf{1}^T_{[p_2]}\}$$

$$Cov(\hat{\boldsymbol{\nu}}_{(p_2)k}) = \frac{\sigma^2 N_k}{(N_k - N^{(2)}_k)N^{(2)}_k}\{(1-\rho)\mathbf{I}_{[p_2]}+[\rho+p_1a_1[1+(p_1-1)\rho]-2p_1a_2\rho]\mathbf{1}_{[p_2]}\mathbf{1}^T_{[p_2]}\}$$

and

$$Cov(\hat{\boldsymbol{\nu}}_{(p_2)k}, \hat{\boldsymbol{\mu}}_{(p_2)k}) = -\frac{\sigma^2}{N^{(2)}_k}\{(1-\rho)\mathbf{I}_{[p_2]}+[\rho+p_1a_1[1+(p_1-1)\rho]-2p_1a_2\rho]\mathbf{1}_{[p_2]}\mathbf{1}^T_{[p_2]}\},$$

where $a_2 = \rho/[1+(p_1-1)\rho]$ *and* $a_1 = [1+(p_1-1)\rho]^{-4}V(\hat{\rho}) + a_2^2$.

Here, $Cov(\hat{\boldsymbol{\mu}}_{(p_2)k}$ has the same expression as (9b) in Carriere (1994a) except for $V(\hat{\rho})$, which is as in (4.4.19).

4.4.2 When $\mathbf{e} \neq \mathbf{0}$

For the case of $\mathbf{e} \neq \mathbf{0}$, we give the results only for $L = 2$. Denote the covariance matrix for the proxy, given the observed data in the first p_1 periods, as $\tilde{\boldsymbol{\Sigma}}_{2.1} = \boldsymbol{\Sigma}_{2.1} + \boldsymbol{\Sigma}_{e22}$, where $\boldsymbol{\Sigma}_{e22}$ is the $p_2 \times p_2$ submatrix of $\boldsymbol{\Sigma}_e$. The component L_2 (4.4.3) becomes

$$\begin{aligned}
L_2 \;\propto\;& |\boldsymbol{\Sigma}_{2.1}|^{-N^{(2)}/2} exp\{-\frac{1}{2}\sum_k \sum_{j=1}^{N_k^{(2)}} [\mathbf{y}_{(p_2)jk} - \boldsymbol{\mu}_{(p_2)k} - \mathbf{B}^{(2)}(\mathbf{y}_{(p_1)jk} - \boldsymbol{\mu}_{(p_1)k})]^T \boldsymbol{\Sigma}_{2.1}^{-1} \\
\times\;& [\mathbf{y}_{(p_2)jk} - \boldsymbol{\mu}_{(p_2)k} - \mathbf{B}^{(2)}(\mathbf{y}_{(p_1)jk} - \boldsymbol{\mu}_{(p_1)k})]\} \\
\times\;& |\tilde{\boldsymbol{\Sigma}}_{2.1}|^{-(N-N^{(2)})/2} exp\{-\frac{1}{2}\sum_k \sum_{j=N_k^{(2)}+1}^{N_k} [\mathbf{p}_{(p_2)jk} - \boldsymbol{\mu}_{(p_2)k} - \boldsymbol{\nu}_{(p_2)k} \\
-\;& \mathbf{B}^{(2)}(\mathbf{y}_{(p_1)jk} - \boldsymbol{\mu}_{(p_1)k})]^T \tilde{\boldsymbol{\Sigma}}_{2.1}^{-1} \\
\times\;& [\mathbf{p}_{(p_2)jk} - \boldsymbol{\mu}_{(p_2)k} - \boldsymbol{\nu}_{(p_2)k} - \mathbf{B}^{(2)}(\mathbf{y}_{(p_1)jk} - \boldsymbol{\mu}_{(p_1)k})]\}.
\end{aligned}$$

Then the MLEs for $\boldsymbol{\mu}_{(p_1)k}$, $\boldsymbol{\mu}_{(p_2)k}$ and $\boldsymbol{\nu}_{(p_2)k}$ are the same as those when $\mathbf{e} = \mathbf{0}$ and the covariance matrices are unchanged, if $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{e22}$ are known.

When $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{e22}$ are unknown, the MLE for $\boldsymbol{\mu}_{(p_2)k}$ and $\boldsymbol{\nu}_{(p_2)k}$ are obtained similarly, using the MLE for $\mathbf{B}^{(2)}$ as

$$\hat{\mathbf{B}}^{(2)} = \mathbf{SS}_{21}^{(2)}(\mathbf{SS}_{11}^{(2)})^{-1} + \hat{\boldsymbol{\Sigma}}_{2.1}\hat{\tilde{\boldsymbol{\Sigma}}}_{2.1}^{-1}\mathbf{SS}_{21}^{(0)}(\mathbf{SS}_{11}^{(2)})^{-1} - \hat{\boldsymbol{\Sigma}}_{2.1}\hat{\tilde{\boldsymbol{\Sigma}}}_{2.1}^{-1}\hat{\mathbf{B}}^{(2)}\mathbf{SS}_{11}^{(0)}(\mathbf{SS}_{11}^{(2)})^{-1},$$

where $\hat{\boldsymbol{\Sigma}}_{2.1} = [\mathbf{SS}_{22}^{(2)} - \hat{\mathbf{B}}^{(2)}\mathbf{SS}_{12}^{(2)} - \mathbf{SS}_{21}^{(2)}(\hat{\mathbf{B}}^{(2)})^T + \hat{\mathbf{B}}^{(2)}\mathbf{SS}_{11}^{(2)}(\hat{\mathbf{B}}^{(2)})^T]/N^{(2)}$ and $\hat{\tilde{\boldsymbol{\Sigma}}}_{2.1} = [\mathbf{SS}_{22}^{(0)} - \hat{\mathbf{B}}^{(2)}\mathbf{SS}_{12}^{(0)} - \mathbf{SS}_{21}^{(0)}(\hat{\mathbf{B}}^{(2)})^T + \hat{\mathbf{B}}^{(2)}\mathbf{SS}_{11}^{(0)}(\hat{\mathbf{B}}^{(2)})^T]/(N - N^{(2)})$ are the MLEs for $\boldsymbol{\Sigma}$ and $\tilde{\boldsymbol{\Sigma}}$, respectively. A solution for $\hat{\mathbf{B}}^{(2)}$ can be obtained iteratively by

$$\hat{\mathbf{B}}_{m+1}^{(2)} = \mathbf{SS}_{21}^{(2)}(\mathbf{SS}_{11}^{(2)})^{-1} + \hat{\boldsymbol{\Sigma}}_{2.1}^{m}(\hat{\tilde{\boldsymbol{\Sigma}}}_{2.1}^{m})^{-1}\mathbf{SS}_{21}^{(0)}(\mathbf{S}_{11}^{(2)})^{-1} - \hat{\boldsymbol{\Sigma}}_{2.1}^{m}(\hat{\tilde{\boldsymbol{\Sigma}}}_{2.1}^{m})^{-1}\hat{\mathbf{B}}_{m}^{(2)}\mathbf{SS}_{11}^{(0)}(\mathbf{SS}_{11}^{(2)})^{-1}$$

until it converges, starting with an initial value $\hat{\mathbf{B}}_{0}^{(2)} = \mathbf{SS}_{21}^{(2)}(\mathbf{SS}_{11}^{(2)})^{-1}$. Here, we denote $\hat{\boldsymbol{\Sigma}}_{2.1}^{m}$ and $\hat{\tilde{\boldsymbol{\Sigma}}}_{2.1}^{m}$ as $\hat{\boldsymbol{\Sigma}}_{2.1}$ and $\tilde{\hat{\boldsymbol{\Sigma}}}_{2.1}$ evaluated at $\hat{\mathbf{B}}_{m}^{(2)}$, respectively. Such a solution $\hat{\mathbf{B}}^{(2)}$ has an asymptotic multivariate normal distribution with mean $\mathbf{B}^{(2)}$ and covariance matrix $\mathbf{I}(\mathbf{B}^{(2)})^{-1}$, where $\mathbf{I}(\mathbf{B}^{(2)})$ is the information matrix for $\mathbf{B}^{(2)}$. Using Corollary 3.2.1 (McDonald and Swaminathan, 1973), we have

$$\begin{aligned}
\mathbf{I}(\mathbf{B}^{(2)}) &= -E(\frac{\partial^2 log(\mathcal{L})}{\partial\mathbf{B}^{(2)}\partial{\mathbf{B}^{(2)}}^T}) \\
&= E[\boldsymbol{\Sigma}_{2.1}^{-1} \bigotimes \mathbf{SS}_{11}^{(2)} + \tilde{\boldsymbol{\Sigma}}_{2.1}^{-1} \bigotimes \mathbf{SS}_{11}^{(0)}] \\
&= \boldsymbol{\Sigma}_{2.1}^{-1} \bigotimes E(\mathbf{SS}_{11}^{(2)}) + \tilde{\boldsymbol{\Sigma}}_{2.1}^{-1} \bigotimes E(\mathbf{SS}_{11}^{(0)}) \\
&= \boldsymbol{\Sigma}_{2.1}^{-1} \bigotimes (N^{(2)} - s)\boldsymbol{\Sigma}_{11} + \tilde{\boldsymbol{\Sigma}}_{2.1}^{-1} \bigotimes (N - N^{(2)} - s)\boldsymbol{\Sigma}_{11} \\
&= [(N^{(2)} - s)\boldsymbol{\Sigma}_{2.1}^{-1} + (N - N^{(2)} - s)\tilde{\boldsymbol{\Sigma}}_{2.1}^{-1}] \bigotimes \boldsymbol{\Sigma}_{11} \\
&= \mathbf{W} \bigotimes \boldsymbol{\Sigma}_{11},
\end{aligned}$$

where $\mathbf{W} = (N^{(2)} - s)\boldsymbol{\Sigma}_{2.1}^{-1} + (N - N^{(2)} - s)\tilde{\boldsymbol{\Sigma}}_{2.1}^{-1}$ and $I^{-1}(\mathbf{B}^{(2)}) = \mathbf{W}^{-1} \otimes \boldsymbol{\Sigma}_{11}^{-1}$.

Thus, the covariance matrix for $\hat{\boldsymbol{\mu}}_{(p_2)k}$ is

$$\begin{aligned} Cov(\hat{\boldsymbol{\mu}}_{(p_2)k}) &= Cov(\overline{\mathbf{y}}^{(2)}_{(p_2).k} - \hat{\mathbf{B}}^{(2)}(\overline{\mathbf{y}}^{(2)}_{(p_1).k} - \overline{\mathbf{y}}^{(1)}_{(p_1).k})) \\ &= E_{\hat{\mathbf{B}}^{(2)}} Cov(\hat{\boldsymbol{\mu}}_{(p_2)k} | \hat{\mathbf{B}}^{(2)}) \\ &= \frac{\boldsymbol{\Sigma}_{22}}{N_k^{(2)}} + \frac{N_k - N_k^{(2)}}{N_k N_k^{(2)}} E_{\hat{\mathbf{B}}^{(2)}}[\hat{\mathbf{B}}^{(2)} \boldsymbol{\Sigma}_{11} (\hat{\mathbf{B}}^{(2)})^T - 2\boldsymbol{\Sigma}_{21} E(\hat{\mathbf{B}}^{(2)})^T] \\ &= \frac{\boldsymbol{\Sigma}_{22}}{N_k^{(2)}} + \frac{N_k - N_k^{(2)}}{N_k N_k^{(2)}} [\sum_{jk} (\boldsymbol{\Sigma}_{11})_{jk} Cov(\hat{\mathbf{B}}_j^{(2)}, \hat{\mathbf{B}}_k^{(2)}) - \mathbf{B}^{(2)} \boldsymbol{\Sigma}_{11} (\mathbf{B}^{(2)})^T] \\ &\approx \frac{\boldsymbol{\Sigma}_{22}}{N_k^{(2)}} + \frac{N_k - N_k^{(2)}}{N_k N_k^{(2)}} [\sum_{i=1}^{p_1} (\mathbf{e}_i^T \boldsymbol{\Sigma}_{11} \bigotimes \mathbf{I}_{[p_2]})(\mathbf{W}^{-1} \bigotimes \boldsymbol{\Sigma}_{11}^{-1})(\mathbf{e}_i \bigotimes \mathbf{I}_{[p_2]}) \\ &\quad -\boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}], \end{aligned} \tag{4.4.20}$$

where $\hat{\mathbf{B}}_j^{(2)}$ is the j^{th} column of $\hat{\mathbf{B}}^{(2)}$, and $\mathbf{e}_i$ is a $p_1 \times 1$ vector of 0 except the i^{th} element, which is 1.

Similarly, the covariance matrix for $\hat{\boldsymbol{\nu}}_{(p_2)k}$ is

$$\begin{aligned} Cov(\hat{\boldsymbol{\nu}}_{(p_2)k}) &\approx \frac{\boldsymbol{\Sigma}_{e22} - \boldsymbol{\Sigma}_{22}}{N_k - N_k^{(2)}} + \frac{N_k}{N_k^{(2)}(N_k - N_k^{(2)})} \\ &\quad \times [\sum_{i=1}^{p_1} (\mathbf{e}_i^T \boldsymbol{\Sigma}_{11} \bigotimes \mathbf{I}_{[p_2]})(\mathbf{W}^{-1} \bigotimes \boldsymbol{\Sigma}_{11}^{-1})(\mathbf{e}_i \bigotimes \mathbf{I}_{[p_2]}) + \boldsymbol{\Sigma}_{2.1}] \end{aligned} \tag{4.4.21}$$

and

$$Cov(\hat{\boldsymbol{\mu}}_{(p_2)k}, \hat{\boldsymbol{\nu}}_{(p_2)k}) \approx -\frac{1}{N_k^{(2)}}[\boldsymbol{\Sigma}_{2.1} + \sum_{i=1}^{p_1} (\mathbf{e}_i^T \boldsymbol{\Sigma}_{11} \bigotimes \mathbf{I}_{[p_2]})(\mathbf{W}^{-1} \bigotimes \boldsymbol{\Sigma}_{11}^{-1})(\mathbf{e}_i \bigotimes \mathbf{I}_{[p_2]})]. \tag{4.4.22}$$

Remark 4.4.5 *When $p_2 = 1$, the covariances (4.4.20), (4.4.21) and (4.4.22) become*

$$\begin{aligned} Cov(\hat{\boldsymbol{\mu}}_{(p_2)k}) &\approx \frac{\sigma_{22}}{N_k^{(2)}} + \frac{N_k - N_k^{(2)}}{N_k N_k^{(2)}}[\frac{p_1 \boldsymbol{\sigma}_{2.1}\tilde{\boldsymbol{\sigma}}_{2.1}}{(N^{(2)} - s)\tilde{\boldsymbol{\sigma}}_{2.1} + (N - N^{(2)} - s)\boldsymbol{\sigma}_{2.1}} \\ &\quad -\boldsymbol{\sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\sigma}_{12}], \end{aligned}$$

$$\begin{aligned} Cov(\hat{\boldsymbol{\nu}}_{(p_2)k}) &\approx \frac{\sigma_{e22} - \sigma_{22}}{N_k - N_k^{(2)}} + \frac{N_k}{N_k^{(2)}(N_k - N_k^{(2)})} \\ &\quad \times[\frac{p_1 \boldsymbol{\sigma}_{2.1}\tilde{\boldsymbol{\sigma}}_{2.1}}{(N^{(2)} - s)\tilde{\boldsymbol{\sigma}}_{2.1} + (N - N^{(2)} - s)\boldsymbol{\sigma}_{2.1}} + \boldsymbol{\sigma}_{2.1}] \end{aligned}$$

and

$$Cov(\hat{\boldsymbol{\mu}}_{(p_2)k}, \hat{\boldsymbol{\nu}}_{(p_2)k}) \approx -\frac{1}{N_k^{(2)}}[\sigma_{2.1} + \frac{p_1 \boldsymbol{\sigma}_{2.1}\tilde{\boldsymbol{\sigma}}_{2.1}}{(N^{(2)} - s)\tilde{\boldsymbol{\sigma}}_{2.1} + (N - N^{(2)} - s)\boldsymbol{\sigma}_{2.1}}],$$

where $\sigma_{2.1} = \sigma_{22} - \boldsymbol{\sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\sigma}_{12}$ and $\tilde{\sigma}_{2.1} = \sigma_{22} + \sigma_{e22} - \boldsymbol{\sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\sigma}_{12}$.

Theorem 4.4.4 *When $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{e22}$ are compound symmetric, i.e., $\boldsymbol{\Sigma} = \sigma^2[(1-\rho)\mathbf{I}_{[p]} + \rho\mathbf{1}_{[p]}\mathbf{1}_{[p]}^T]$ and $\boldsymbol{\Sigma}_{e22} + \boldsymbol{\Sigma}_{22} = \sigma_a^2[(1-\rho_a)\mathbf{I}_{[p_2]} + \rho_a\mathbf{1}_{[p_2]}\mathbf{1}_{[p_2]}^T]$, the MLEs for $\boldsymbol{\mu}_{(p_2)k}$ and $\boldsymbol{\nu}_{(p_2)k}$ are same as in Remark 4.4.3 with $\hat{\rho}$, an approximately unbiased estimator for ρ, obtained as*

$$\hat{\rho} = \frac{2\sum_{i>i'}^{p}(\mathbf{SS}^{(2)})_{ii'}}{(p-1)\sum_{1}^{p}(\mathbf{SS}^{(2)})_{ii}}$$

and the variance for $\hat{\rho}$ is

$$V(\hat{\rho}) = \frac{2(1-\rho)^2[1+\rho(p-1)]^2}{p(p-1)(N^{(2)}-s)}. \tag{4.4.23}$$

The unbiased estimator for σ^2 is

$$\hat{\sigma}^2 = \frac{\sum_1^p(\mathbf{SS}^{(2)})_{ii}}{p(N^{(2)}-s)},$$

and the variance for $\hat{\sigma}^2$ is

$$V(\hat{\sigma}^2) = 2\sigma^4\frac{1+\rho^2(p-1)}{p(N^{(2)}-s)},$$

as in Carriere (1994a). The unbiased estimator for σ_a^2 is

$$\hat{\sigma}_a^2 = \frac{\sum_{i=p_1+1}^p(\mathbf{SS}_{22}^{(0)})_{ii}}{p_2(N-N^{(2)}-s)},$$

and the variance for $\hat{\sigma_a}^2$ is

$$V(\hat{\sigma}_a^2) = 2\sigma_a^4\frac{1+\rho_a^2(p_2-1)}{p_2(N-N^{(2)}-s)}.$$

An approximately unbiased estimator for ρ_a is obtained as

$$\hat{\rho}_a = \frac{2\sum_{i>i'=p_1+1}^{p}(\mathbf{SS}_{22}^{(0)})_{ii'}}{(p_2-1)\sum_{i=p_1+1}^{p}(\mathbf{SS}_{22}^{(0)})_{ii}},$$

and its variance is

$$V(\hat{\rho}_a) = \frac{2(1-\rho_a)^2[1+\rho_a(p_2-1)]^2}{p_2(p_2-1)(N-N^{(2)}-s)}.$$

PROOF. Since the proxy and the observation have different covariance matrices for the second group of periods, we used the complete subset data to estimate σ^2 and ρ as in Carriere (1994a). Then, we use the proxy information to estimate σ_a^2 and ρ_a. Theorem 4.4.4 is proved following the similar derivation of Theorem 4.4.3.

□

The covariance $Cov(\hat{\boldsymbol{\mu}}_{(p_2)k})$ is the same as (9b) in Carriere (1994a), $Cov(\hat{\boldsymbol{\nu}}_{(p_2)k})$ is the same as in Remark 4.4.4 with an addition $\{\sigma_a^2[(1-\rho_a)\mathbf{I}_{[p_2]}+\rho_a\mathbf{1}_{[p_2]}\mathbf{1}_{[p_2]}^T] - \sigma^2[(1-\rho)\mathbf{I}_{[p_2]}+\rho\mathbf{1}_{[p_2]}\mathbf{1}_{[p_2]}^T]\}/(N_k-N_k^{(2)})$ and $V(\hat{\rho})$ as given in (4.4.23). The covariance $Cov(\hat{\boldsymbol{\mu}}_{(p_2)k}, \hat{\boldsymbol{\nu}}_{(p_2)k})$ is the same as that in Remark 4.4.4 with $V(\hat{\rho})$ as given in (4.4.23).

4.5 Testing procedures

The hypotheses we are interested in testing are that there are no direct treatment effects ($H_{01} : \tau = 0$) and no residual effects ($H_{02} : \gamma = 0$). The test statistic we consider is of the form

$$t_\theta = \frac{\hat{\theta}}{\sqrt{v\hat{a}r(\hat{\theta})}},$$

where $v\hat{a}r(\hat{\theta})$ is the estimated variance for estimator $\hat{\theta}$, and $\hat{\theta}$ is either $\hat{\tau}$ or $\hat{\gamma}$. Under the assumption of normality, the statistic t_θ has an approximate t-distribution with certain degrees of freedom. Carriere (1994a and 1999) and Patel (1991) chose the following degrees of freedom for testing H_{01} and H_{02}. For compound symmetric covariance structures, the degrees of freedom are $(p-1)(N^{(L)} - s)$ for testing both τ and γ. When no pattern can be specified for the covariance matrix, the degrees of freedom for τ and γ were proposed as $N^{(L)} - s$ and $(N^{(L)} + N - 2s - p_1 p_2)/2$, respectively.

When $\mathbf{e} = \mathbf{0}$, for the approach utilizing proxy information, we propose the degrees of freedom $d_1 = (p-1)(N-2s)$ under the compound symmetric covariance model. Note that d_1 is the degrees of freedom associated with the sum of squared error SSE defined in Theorem 4.4.3. When the covariance matrix is unknown and no pattern could be specified, we follow the suggestion of others, such as Patel (1991) and Carriere (1999), and propose the degrees of freedom of $d_2 = N - (3s + p_1)/2$, where d_2 is the average of the degrees of freedom associated with Σ_{11} and $\Sigma_{2.1}$ (see Section 4.4). The proposed degrees of freedom work well, as we will show in the next section.

When $\mathbf{e} \neq \mathbf{0}$, recall that the extra proxy information does not contribute to the MLEs for $\boldsymbol{\mu}$ as shown in Theorem 4.4.4, in the case where both $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{e22}$ are compound symmetric. For this reason, we use the same degrees of freedom $(p-1)(N^{(2)} - s)$ as in the incomplete data method (Carriere, 1994a). When $\mathbf{e} \neq \mathbf{0}$ and no specific covariance form can be used, we apply the average of the degrees of freedom associated with $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{2.1}$ and $\tilde{\boldsymbol{\Sigma}}_{2.1}$, ignoring the uncertainty of $\mathbf{B}^{(2)}$, which is $d_3 = (2N - 3s - 2p_1)/3$ (see Section 4.4).

4.6 Simulation study

We compare the performance of the estimators $\hat{\tau}$ and $\hat{\gamma}$ in the two approaches mentioned above, using small samples. We generated 1000 samples from model (4.2.1) with $\boldsymbol{\beta} = \mathbf{0}$ and $\boldsymbol{\eta} = .\mathbf{5}$ for Design I and Design IV (defined in Chapter 2), with 5 or 10 subjects in each sequence. We consider the case where 20% of the data obtained is proxy data and the missing data pattern is monotonic.

When $\mathbf{e} \neq \mathbf{0}$, the sample size is chosen to be 10 for each sequence. The covariance matrix $\mathbf{\Sigma}$ (in model (4.2.1)) that we considered is

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & \rho\sqrt{c_1} & \rho\sqrt{c_2} \\ & c_1 & \rho\sqrt{c_1 c_2} \\ & & c_2 \end{bmatrix}$$

with two different values of ρ, which were .3 and .7, and two different values for c_1 and c_2, which are 1 and 4. The covariance matrix for the data with the proxy component is set at $c_3\mathbf{\Sigma}$, where $c_3 = 4$. For Design I, the last column and row are deleted from $\mathbf{\Sigma}$. The empirical sizes and powers of the test against $H_{11} : \theta = .5$ are calculated from both approaches. For Design IV, we consider testing against $H_{11} : \theta = .25$.

Table 4.1 reports the size and power of the test for H_{01} for Design I with sample sizes 10 and 20, when $\mathbf{e} = \mathbf{0}$. The suggested degrees of freedom appear to work well in testing the direct treatment effects and are shown to be more powerful than the incomplete data analysis methods without utilizing proxies. Table 4.2 shows the power comparison for testing γ for Design I. The suggested degrees of freedom appear to keep the nominal level satisfactorily. However, the testing procedure for the residual treatment effect using proxy has a similar power to that of traditional incomplete data analysis, although the use of proxy appears to be less efficient when the sample size is small and the covariance matrix is unspecified.

Tables 4.3-4.6 tabulate similar results for Design IV. For testing τ, the procedure that uses proxy is more powerful except in the case of relatively large sample size ($n = 20$) and a ratio of variances of .3, where the two approaches perform similarly. For testing γ, the procedure using proxy is better, even for a small sample size ($n = 10$) and high ratio of variances ($\rho = 0.7$). The two approaches perform similarly in the other cases. In general, the method using proxy appears to be more powerful than using the traditional incomplete data analysis strategy.

When $\mathbf{e} \neq \mathbf{0}$, the proxy method and the incomplete method performed similarly, when the covariance matrix is compound symmetric. Under an unspecified covariance matrix, we found that the suggested degrees of freedom work well for both τ and γ. For Design I, Table 4.7 demonstrates that the incomplete method is more powerful than the proxy method for testing H_{01}, when ρ is relatively small and the missing proportion is relatively small, i.e., 20%. For testing H_{02}, the two approaches perform similarly (see Table 4.8).

For Design IV, tables 4.9 and 4.10 demonstrate that the incomplete method is more powerful than the proxy method for testing H_{01} and H_{02}.

Although not shown here, when the proxy is used to fill in 50% of the missing data, the proxy approach is more powerful than the conventional incomplete data strategy.

We also investigated a relatively larger bias in proxy with $\boldsymbol{\eta} = 1$, and a smaller discrepancy between the covariance matrices for observed data and proxy (i.e., $c_3 = 2$). Our findings remain unchanged; overall, the procedure using proxy is more powerful than the conventional incomplete data method when $\mathbf{e} = \mathbf{0}$. When $\mathbf{e} \neq \mathbf{0}$, the proxy method shows its superiority for testing H_{01} when the missing proportion is relatively large. For testing H_{02}, generally speaking, the proxy method is just as powerful as the conventional incomplete data method when the proxy has a different covariance matrix from the observed actual data.

In summmary, we find that the use of proxy data to replace missing values is an excellent alternative for missing data problems. Further, it allows the researchers to use standard analysis software. In principle, this approach (with proxy) is equivalent to the single imputation method. In the next chapter, we contrast this approach with the multiple imputation method.

4.7 Design considerations

We have compared two alternative strategies for handling incomplete data and demonstrated that the use of proxies is generally efficient. We now consider what proportion of proxy data can be introduced to achieve a given efficiency. We discuss design considerations here because ultimately we want the chosen design to be efficient. We will derive the conditions under which the use of proxy data can be nearly as efficient as when all data are observed. Consider that we wish to test

$$H_0 : a^T \boldsymbol{\beta} = a_0, \tag{4.7.1}$$

where a is a $q \times 1$ vector containing 0's and 1's, and a_0 is the value of $a^T\boldsymbol{\beta}$ specified under the null hypothesis. Then, $a^T\boldsymbol{\beta}$ is estimated by $a^T\hat{\boldsymbol{\beta}}$ and its variance by $a^T\mathbf{I}(\boldsymbol{\beta})^{-1}a$. If all data are actual, these are estimated by $a^T\hat{\boldsymbol{\beta}}^f$ and $a^T\mathbf{I}(\boldsymbol{\beta}^f)^{-1}a$.

Definition 4.7.1 *The estimator $a^T\hat{\boldsymbol{\beta}}$ using proxy data is at least $c \times 100\%$ as efficient as $a^T\hat{\boldsymbol{\beta}}^f$ if*

$$\frac{a^T\mathbf{I}(\boldsymbol{\beta}^{\mathbf{f}})^{-1}\mathbf{a}}{a^T\mathbf{I}(\boldsymbol{\beta})^{-1}\mathbf{a}} \le c, \tag{4.7.2}$$

where c is a real number between 0 and 1.

Definition (4.7.1) can be used to specify the conditions for $\delta_{i.k}$, the maximum number of proxy (i.e., missing observations) for period i and sequence k that can be allowed to attain $c \times 100\%$ efficiency.

It is difficult to make general statements, but some special cases are treated in Sections 4.7.1 and 4.7.2. For two-treatment trials, we consider the *dual-balanced* design as in Carriere and Huang (2000, 2001). Dual-balanced designs allocate an equal number of subjects to sequence k and its dual k^*, where the treatments assigned for sequence k are the opposite of those for the sequence k^* for each period.

In sections 4.7.1 and 4.7.2, we obtain conditions for $\delta_{i.k}$ for some special designs to produce estimators for τ that are at least $c \times 100\%$ efficient. Here, we do not include the sequence effects in model (4.2.1), and we assume that the pattern of missing data (which were filled in by their proxy) is monotonic and the data in the first period is complete for all subjects. Also, to simplify the discussion, we assume that the proxy and the actual data are similar in their variability ($\mathbf{e} = \mathbf{0}$). The case of $\delta_{i.k} = N_k$ for any $i > 1$ can be considered as the case of pseudo repeated measures design where an extra period(s) is created using proxy information in the complete randomized design.

4.7.1 Two-period two-treatment designs

For two-period designs, we consider Designs I, II and III (as defined in Chapter 2, Table 2.5) to investigate the use of proxy in design considerations.

Let $r_1 = \delta_{2.1}/(N/4) = \delta_{2.2}/(N/4)$ and $r_2 = \delta_{2.3}/(N/4) = \delta_{2.4}/(N/4)$ be the proportions of data to be filled in by proxies in the second period of sequences AB, BA, AA and BB, respectively, in a dual-balanced design. We obtain conditions for $\delta_{2.k}$, $k = 1, \ldots, 4$, to produce at least $c \times 100\%$ efficient estimators for τ for two-period designs.

Result 4.7.1 *For Design I and Design III, the estimator for the treatment effect contrast τ using proxy information in the second period is as efficient as* *the estimator based on complete actual data, regardless of the fraction of proxy data and the measurement error structure under the model (4.2.1).*

Result 4.7.1 is due to the fact that, in these two-period designs, the estimator for τ uses only the first period data under the model, with unequal residual effects. Hence, use of proxy information to fill in the missing second period data does not result in any gain or loss in efficiency.

Result 4.7.2 *For Design II, the estimator for the treatment effect contrast* τ *using proxy information can be at least* $c \times 100\%$ *as efficient as the estimator using complete actual data, if the proportions* r_1 *and* r_2 *are chosen such that*

$$r_1 \leq \frac{-b_1 r_2 - b_2}{r_2 + b_1} \tag{4.7.3}$$

and

$$r_2 \leq -b_2/b_1, \tag{4.7.4}$$

when $\mathbf{\Sigma} = \sigma^2[(1-\rho)\mathbf{I}_{[p]} + \rho\mathbf{1}_{[p]}\mathbf{1}_{[p]}^T]$*, where* $b_1 = [-2b^2 + 6b - 3 + (b^2 - 4b + 2)c]/[2(1-b)^2]$ *and* $b_2 = [(b^2 - 4b + 2)(1-c)]/(1-b)^2$*, with* $b = \rho/(1+\rho)$*. We see that they reduce to* $r_1 \leq [4(c-1) - (2c-3)r_2]/(2r_2 + 2c - 3)$ *and* $r_2 \leq 4(c-1)/(2c-3)$ *when* $\Sigma = \sigma_\varepsilon^2\mathbf{I}_{[p]}$.

Table 4.11 reports the possible proportions of proxy information permitted to generate a nearly efficient estimator for τ for Design II as defined by $c > .8$.

As expected, efficiency increases as the proxy proportion decreases. For a given efficiency, high within-subject correlation has an implication that not as much proxy is allowed as in the case of a low value of ρ. The proportions of proxy information permitted in the sequence AA(BB) increase as those in the sequence AB(BA) decrease.

This suggests that more resources should be allocated to collect actual information in sequences AB(BA), as sequences AA(BB) are less desirable for clinicians (Carriere and Reinsel, 1992). For example, when $\rho = 0$, to obtain a 90% efficient estimator for τ, up to 20% of the data in the second period of the sequences AA(BB) and 20% of the data in the sequences AB(BA) can be collected from proxies. When $\rho = 0.6$, we could have 20% proxy data in the second period of the sequences AA(BB) and 12% proxy data in the sequences AB(BA), with less than 10% loss of efficiency for estimating τ.

4.7.2 Three-period two-treatment designs

For three-period designs, we will consider Design IV and Design VI, which, in Chapter 2, were noted to be the most robust. To simplify the discussion, we assume that the missing data occurs only in the third (last) period.

Let $r_1 = \delta_{3.1}/(N/4) = \delta_{3.2}/(N/4)$ and $r_2 = \delta_{3.3}/(N/4) = \delta_{3.4}/(N/4)$ be the proportions of data filled in by proxy data in the third period of the sequences ABB(BAA) and AAB(BBA), respectively. We have the following results for three-period designs.

Result 4.7.3 *For Design IV, the estimator for the treatment effect contrast* τ *using the proxy information can be at least* $c \times 100\%$ *as efficient as the estimator using complete actual data if the proportion* r_1 *is chosen such that*

$$r_1 \leq \frac{2b^2 - 8b + 6 + 2c(3-b)(b-1)}{2b^2 - 5b + 5 - c(3-b)} \tag{4.7.5}$$

when $\mathbf{\Sigma} = \sigma^2[(1-\rho)\mathbf{I}_{[p]} + \rho \mathbf{1}_{[p]}\mathbf{1}_{[p]}^T]$, *where* $b = \rho/(1+2\rho)$. *This reduces to* $r_1 \leq 6(1-c)/(5-3c)$ *when* $\mathbf{\Sigma} = \sigma_\varepsilon^2 \mathbf{I}_{[p]}$.

For Design IV, Table 4.12 shows that the proportion of proxy information permitted in the third period is decreasing with increasing correlation coefficient ρ and increasing c. Unlike two-period designs, the proportion of possible proxy information in the third period is an increasing function in ρ.

Result 4.7.4 *For Design VI, to attain at least* $c \times 100\%$ *efficiency in estimating* τ, *the proportions* r_1 *and* r_2 *must be*

$$r_1 \leq \frac{-a_3 - a_2 r_2}{a_0 r_2 + a_1} \tag{4.7.6}$$

and

$$r_2 \leq min(-a_1/a_0, -a_3/a_2), \tag{4.7.7}$$

when $\mathbf{\Sigma} = \sigma^2[(1-\rho)\mathbf{I}_{[p]} + \rho\mathbf{1}_{[p]}\mathbf{1}_{[p]}^T]$, *where* $a_0 = -2(2b^4 - 4b^3 + 6b^2 - 4b + 2), a_1 = -2(-2b^4 + 6b^3 - 9b^2 + 10b - 5) - c(b^2 - 8b + 6), a_2 = -2(-2b^4 + 18b^3 - 29b^2 + 18b - 5) - c(b^2 - 8b + 6)(4b^2 - 4b + 1)$, *and* $a_3 = -2(2b^4 - 20b^3 + 46b^2 - 40b + 12) - c(b^2 - 8b + 6)(-4b^2 + 8b - 4)$. *This reduces to* $r_1 \leq [-24(1-c) + (10-6c)r_2]/[4r_2 - (10-6c)]$ *and* $r_2 \leq 6(1-c)/(5-3c)$ *when* $\Sigma = \sigma_\varepsilon^2\mathbf{I}_{[p]}$.

The results shown in Table 4.13 for Design VI are similar to those in Table 4.11 for Design II. If the sequences ABB and BAA are preferable to the sequences AAB and BBA, more resources should be allocated to collecting as much actual data as possible from the sequences ABB and BAA. For example, when $\mathbf{\Sigma} = \sigma_\varepsilon^2\mathbf{I}_{[p]}$, we could use 50% and 68% proxy data in the third period of the sequences AAB, BBA, ABB and BAA, respectively, with just 20% loss in efficiency.

4.8 Discussion

We have investigated the role of proxy information in statistical analyses. Proxy information can be available in diverse forms. Its quality also varies a great deal. For example, unavailable information about one's socio-economic status can be filled in with proxy data on neighborhood affluence. In a clinical trial context, the patients' care providers can provide pertinent information.

The strategy of collecting proxy information in an effort to optimize the data collection process is particularly useful in statistical data analyses. In essence, this approach is similar to the single imputation strategy, where missing data are filled in with proxy values generated via various methods (Little and Rubin, 1987). Our consideration, however, would be more informative and personal than the commonly used single imputation. Common software gives the users a choice of the method of imputation to adopt. The most popular imputation method may be to use the mean of a sub-group whose particular covariate pattern is shared by the subjects that have missing data.

What we have learned in this study is rather substantial and significant. Even when over 50% of the data are replaced by proxy information, we find that the efficiency, as compared to the ideal situation of complete actual data, can be over 90%. As well, analysts can use this method while enjoying all the conveniences that are normally available to them.

Comparisons to other existing techniques revealed that estimators utilizing proxy information obtain better results than they would by using the incomplete analytic method, especially when the proxy and the observed data share the same covariance matrix.

Table 4.1: Empirical sizes and powers for testing τ in Design I when $\mathbf{e} = \mathbf{0}$

n	ρ	c_1	Procedure	type	$\alpha = .01$		$\alpha = .05$		$\alpha = .1$	
					size	power	size	power	size	power
10	.3	1	prx	sys	.007	.158	.053	.467	.089	.631
				uns	.007	.182	.052	.480	.094	.629
			inc	sys	.008	.117	.049	.381	.099	.592
				uns	.008	.116	.046	.390	.101	.558
		4	prx	sys	.016	.072	.061	.233	.110	.352
				uns	.014	.065	.052	.213	.112	.322
			inc	sys	.009	.042	.051	.194	.108	.318
				uns	.009	.040	.045	.165	.097	.282
	.7	1	prx	sys	.006	.416	.055	.768	.100	.885
				uns	.009	.439	.060	.761	.110	.875
			inc	sys	.009	.284	.048	.685	.105	.846
				uns	.009	.261	.050	.643	.099	.820
		4	prx	sys	.007	.093	.053	.295	.106	.456
				uns	.010	.095	.047	.290	.099	.444
			inc	sys	.008	.065	.050	.264	.104	.414
				uns	.010	.060	.046	.250	.095	.393
20	.3	1	prx	sys	.008	.580	.048	.840	.099	.921
				uns	.008	.595	.051	.837	.101	.921
			inc	sys	.010	.514	.049	.805	.109	.909
				uns	.012	.513	.050	.794	.108	.906
		4	prx	sys	.015	.225	.062	.422	**.125**	.540
				uns	.007	.173	.048	.383	.103	.500
			inc	sys	.013	.184	.064	.389	.111	.516
				uns	.010	.149	.055	.363	.099	.477
	.7	1	prx	sys	.008	.943	.042	.991	.086	.998
				uns	.010	.944	.043	.993	.086	.997
			inc	sys	.007	.900	.040	.989	.089	.996
				uns	.008	.900	.038	.989	.094	.995
		4	prx	sys	.012	.360	**.071**	.613	**.121**	.742
				uns	.006	.343	.052	.598	.107	.724
			inc	sys	.016	.307	.063	.578	**.129**	.725
				uns	.011	.279	.049	.572	.109	.709

Note:

prx—Approach utilizing proxy data

inc—Incomplete data procedure (Carriere, 1994a and 1999)

sys—Compound symmetry pattern of covariance matrix

uns—Unstructured covariance pattern

Table 4.2: Empirical sizes and powers for testing γ in Design I when $\mathbf{e} = \mathbf{0}$

n	ρ	c_1	Procedure	type	$\alpha = .01$		$\alpha = .05$		$\alpha = .1$	
					size	power	size	power	size	power
10	.3	1	prx	sys	.010	.026	.043	.112	.098	.184
				uns	.010	.030	.048	.106	.105	.185
			inc	sys	.011	.023	.041	.096	.086	.180
				uns	**.018**	.034	.053	.121	.097	.199
		4	prx	sys	.006	.011	.039	.076	.088	.150
				uns	.006	.011	.038	.070	.083	.134
			inc	sys	.011	.016	.040	.070	.083	.131
				uns	.010	.021	.051	.080	.094	.146
	.7	1	prx	sys	.005	.020	.042	.082	.083	.177
				uns	.004	.018	.041	.095	.082	.180
			inc	sys	.007	.013	.040	.075	.097	.161
				uns	.007	.017	.044	.083	.091	.179
		4	prx	sys	.009	.015	.046	.061	.088	.127
				uns	.005	.012	.037	.060	.081	.110
			inc	sys	.007	.009	.044	.063	.091	.124
				uns	.009	.016	.045	.068	.083	.123
20	.3	1	prx	sys	.009	.073	.043	.213	.081	.324
				uns	.010	.078	.041	.216	.083	.318
			inc	sys	.013	.073	.043	.208	.091	.316
				uns	.012	.079	.048	.218	.092	.322
		4	prx	sys	.014	.034	.052	.121	.109	.201
				uns	.011	.027	.043	.099	.098	.183
			inc	sys	.015	.026	.058	.115	.094	.195
				uns	.013	.025	.048	.096	.099	.180
	.7	1	prx	sys	.013	.063	.045	.205	.091	.322
				uns	.012	.067	.048	.208	.096	.312
			inc	sys	.010	.059	.044	.201	.093	.312
				uns	.010	.062	.048	.200	.094	.311
		4	prx	sys	.011	.031	.052	.116	.101	.195
				uns	.011	.023	.043	.096	.086	.175
			inc	sys	.011	.025	.045	.103	.102	.184
				uns	.011	.025	.040	.094	.091	.171

Note:
prx—Approach utilizing proxy
inc—Incomplete data procedure (Carriere, 1994a and 1999)
sys—Compound symmetry pattern of covariance matrix
uns—Unstructured covariance pattern

Table 4.3: Empirical sizes and powers for testing τ in Design IV when $\mathbf{e} = \mathbf{0}$ and $n = 10$

ρ	c_1	Procedure	type	$\alpha = .01$		$\alpha = .05$		$\alpha = .1$	
	c_2			size	power	size	power	size	power
.3	1	prx	sys	.011	.090	.047	.250	.095	.372
	1		uns	.010	.053	.041	.204	**.072**	.333
		inc	sys	.012	.069	.049	.237	.096	.371
			uns	.005	.034	.040	.155	**.073**	.287
	1	prx	sys	.010	.042	**.034**	.152	.086	.249
	4		uns	.008	.028	.050	.141	.094	.242
		inc	sys	.007	.035	**.031**	.139	.081	.236
			uns	.006	.022	.040	.111	.094	.193
	4	prx	sys	.004	.030	**.023**	.136	**.066**	.227
	1		uns	.006	.050	.054	.171	.118	.281
		inc	sys	.006	.027	**.031**	.131	**.073**	.219
			uns	.004	.026	.040	.136	.096	.231
.7	1	prx	sys	.013	.253	**.067**	.512	**.125**	.643
	1		uns	.010	.158	.063	.435	**.120**	.603
		inc	sys	.016	.208	**.067**	.477	**.130**	.630
			uns	.007	.083	.048	.324	.093	.510
	1	prx	sys	.013	.069	.042	.204	.081	.329
	4		uns	.010	.048	.052	.202	.087	.327
		inc	sys	.008	.057	.050	.298	.091	~~.332~~
			uns	.013	.028	.038	.142	**.079**	.279
	4	prx	sys	.004	.060	**.019**	.207	**.051**	.323
	1		uns	.009	.080	.041	.258	.085	.407
		inc	sys	.007	.056	**.026**	.201	**.055**	.319
			uns	.006	.036	**.029**	.198	**.070**	.349

Note:

prx—Approach utilizing proxy data

inc—Incomplete data procedure (Carriere, 1994a and 1999)

sys—Compound symmetry pattern of covariance matrix

uns—Unstructured covariance pattern

Table 4.4: Empirical sizes and powers for testing τ in Design IV when $\mathbf{e} = \mathbf{0}$ and $n = 20$

ρ	c_1	Procedure	type	$\alpha = .01$		$\alpha = .05$		$\alpha = .1$	
	c_2			size	power	size	power	size	power
.3	1	prx	sys	.008	.259	.043	.524	.090	.670
	1		uns	.007	.232	.046	.491	.088	.640
		inc	sys	.008	.256	.051	.519	.090	.663
			uns	.007	.290	.038	.442	.082	.616
	1	prx	sys	.007	.105	.045	.263	.084	.381
	4		uns	.012	.112	.053	.276	.111	.420
		inc	sys	.009	.104	.049	.250	.084	.376
			uns	.012	.092	.052	.256	.103	.402
	4	prx	sys	**.002**	.069	**.025**	.248	**.055**	.385
	1		uns	.009	.130	.048	.360	.096	.492
		inc	sys	**.003**	.079	**.026**	.249	**.059**	.388
			uns	.005	.102	.041	.319	.088	.474
.7	1	prx	sys	.013	.656	.046	.837	.094	.912
	1		uns	.013	.574	.043	.823	.092	.890
		inc	sys	.013	.610	.050	.823	.098	.903
			uns	.011	.480	.041	.782	.089	.890
	1	prx	sys	.010	.213	.045	.446	.090	.581
	4		uns	.010	.250	.065	.515	.119	.649
		inc	sys	.010	.201	.041	.430	.081	.569
			uns	.011	.213	.060	.484	.108	.611
	4	prx	sys	.004	.173	**.018**	.454	**.045**	.604
	1		uns	.013	.335	.049	.614	.106	.730
		inc	sys	.004	.154	**.025**	.427	**.055**	.585
			uns	.010	.261	.044	.567	.096	.699

Note:
prx—Approach utilizing proxy data
inc—Incomplete data procedure (Carriere, 1994a and 1999)
sys—Compound symmetry pattern of covariance matrix
uns—Unstructured covariance pattern

Table 4.5: Empirical sizes and powers for testing γ in Design IV when $\mathbf{e} = \mathbf{0}$ and $n = 10$

ρ	c_1 c_2	Procedure	type	$\alpha = .01$		$\alpha = .05$		$\alpha = .1$	
				size	power	size	power	size	power
.3	1	prx	sys	.015	.054	.054	.169	.092	.274
	1		uns	.008	.032	.049	.145	.089	.246
		inc	sys	.010	.045	.049	.164	.090	.278
			uns	.014	.039	.053	.141	.101	.233
	1	prx	sys	**.032**	.056	**.110**	.149	**.175**	.231
	4		uns	.011	.016	.061	.089	.119	.144
		inc	sys	**.026**	.050	**.102**	.136	**.175**	.224
			uns	.012	.019	.049	.068	.106	.126
	4	prx	sys	**.018**	.044	.064	.128	.117	.199
	1		uns	.010	.022	.043	.090	.087	.164
		inc	sys	**.018**	.039	.062	.130	**.123**	.208
			uns	.010	.028	.057	.089	.107	.159
.7	1	prx	sys	.010	.140	.054	.350	.118	.466
	1		uns	.014	.095	.058	.290	.110	.423
		inc	sys	.011	.112	.049	.336	.105	.461
			uns	.011	.078	.055	.233	.113	.371
	1	prx	sys	**.035**	.082	**.110**	.217	**.172**	.324
	4		uns	.007	.021	.047	.106	.099	.199
		inc	sys	**.038**	.069	**.109**	.212	**.181**	.323
			uns	.012	.024	.049	.101	.097	.175
	4	prx	sys	**.022**	.060	**.068**	.170	**.121**	.262
	1		uns	.013	.028	.050	.126	.098	.213
		inc	sys	**.021**	.047	**.081**	.168	**.131**	.261
			uns	.015	.024	.048	.111	.090	.195

Note:

prx—Approach utilizing proxy data
inc—Incomplete data procedure (Carriere, 1994a and 1999)
sys—Compound symmetry pattern of covariance matrix
uns—Unstructured covariance pattern

Table 4.6: Empirical sizes and powers for testing γ in Design IV when $\mathbf{e} = \mathbf{0}$ and $n = 20$

ρ	c_1	Procedure	type	$\alpha = .01$		$\alpha = .05$		$\alpha = .1$	
	c_2			size	power	size	power	size	power
.3	1	prx	sys	.012	.131	.057	.336	.098	.448
	1		uns	.013	.121	.046	.315	.103	.436
		inc	sys	.011	.130	.058	.328	.102	.454
			uns	.010	.126	.059	.311	.120	.438
	1	prx	sys	**.037**	.092	**.115**	.251	**.178**	.335
	4		uns	.016	.033	.066	.138	.117	.236
		inc	sys	**.033**	.097	**.116**	.236	**.177**	.333
			uns	.020	.042	.061	.149	.124	.228
	4	prx	sys	.011	.066	.059	.182	.110	.287
	1		uns	.009	.045	.046	.149	.096	.272
		inc	sys	.011	.064	.061	.182	.113	.284
			uns	.011	.044	.048	.154	.096	.252
.7	1	prx	sys	.013	.361	.054	.644	.100	.750
	1		uns	.014	.336	.049	.585	.110	.718
		inc	sys	.014	.345	.053	.611	.103	.743
			uns	.012	.322	.066	.584	**.124**	.710
	1	prx	sys	**.046**	.173	**.119**	.326	**.200**	.426
	4		uns	.015	.074	.050	.215	.112	.330
		inc	sys	**.042**	.161	**.117**	.312	**.194**	.420
			uns	.015	.073	.060	.220	.114	.320
	4	prx	sys	.014	.117	.055	.293	.105	.407
	1		uns	.007	.093	.044	.279	.092	.388
		inc	sys	.015	.122	.056	.293	.114	.399
			uns	.007	.092	.050	.271	.094	.384

Note:
prx—Approach utilizing proxy data
inc—Incomplete data procedure (Carriere, 1994a and 1999)
sys—Compound symmetry pattern of covariance matrix
uns—Unstructured covariance pattern

Table 4.7: Empirical sizes and powers for testing τ in Design I when $\mathbf{e} \neq \mathbf{0}$ and $n = 20$

ρ	c_1	Procedure	type	$\alpha = .01$		$\alpha = .05$		$\alpha = .1$	
				size	power	size	power	size	power
.3	1	prx	sys	0.009	0.697	0.058	0.888	0.100	0.953
			uns	0.009	0.647	0.056	0.877	0.100	0.943
		inc	sys	0.009	0.697	0.058	0.888	0.100	0.953
			uns	0.011	0.680	0.060	0.882	0.101	0.951
	4	prx	sys	0.010	0.231	0.050	0.495	0.107	0.647
			uns	0.012	0.201	0.051	0.461	0.106	0.605
		inc	sys	0.010	0.231	0.050	0.495	0.107	0.647
			uns	0.010	0.205	0.043	0.470	0.095	0.619
.7	1	prx	sys	0.008	0.975	0.050	0.996	0.101	0.998
			uns	0.010	0.964	0.050	0.994	0.105	0.996
		inc	sys	0.008	0.975	0.050	0.996	0.101	0.998
			uns	0.007	0.976	0.048	0.996	0.101	0.998
	4	prx	sys	0.007	0.418	0.056	0.678	0.116	0.802
			uns	0.007	0.399	0.052	0.678	0.106	0.802
		inc	sys	0.007	0.418	0.056	0.678	0.116	0.802
			uns	0.007	0.397	0.046	0.670	0.099	0.801

Note:
prx—Approach utilizing proxy data
inc—Incomplete data procedure (Carriere, 1994a and 1999)
sys—Compound symmetry pattern of covariance matrix
uns—Unstructured covariance pattern

Table 4.8: Empirical sizes and powers for testing γ in Design I when $\mathbf{e} \neq \mathbf{0}$ and $n = 20$

ρ	c_1	Procedure	type	$\alpha = .01$		$\alpha = .05$		$\alpha = .1$	
				size	power	size	power	size	power
.3	1	prx	sys	0.012	0.082	0.055	0.222	0.100	0.334
			uns	0.006	0.076	0.055	0.217	0.100	0.334
		inc	sys	0.012	0.082	0.055	0.222	0.100	0.334
			uns	0.010	0.087	0.060	0.215	0.100	0.336
	4	prx	sys	0.009	0.037	0.054	0.141	0.107	0.240
			uns	0.005	0.033	0.049	0.135	0.100	0.229
		inc	sys	0.009	0.037	0.054	0.141	0.107	0.240
			uns	0.007	0.039	0.047	0.135	0.095	0.228
.7	1	prx	sys	0.013	0.085	0.060	0.200	0.108	0.330
			uns	0.009	0.071	0.051	0.200	0.101	0.331
		inc	sys	0.013	0.085	0.060	0.200	0.108	0.330
			uns	0.012	0.082	0.056	0.208	0.103	0.336
	4	prx	sys	0.010	0.033	0.046	0.130	0.093	0.215
			uns	0.006	0.022	0.039	0.126	0.089	0.215
		inc	sys	0.010	0.033	0.046	0.130	0.093	0.215
			uns	0.006	0.031	0.040	0.133	0.085	0.211

Note:
prx—Approach utilizing proxy data
inc—Incomplete data procedure (Carriere, 1994a and 1999)
sys—Compound symmetry pattern of covariance matrix
uns—Unstructured covariance pattern

Table 4.9: Empirical sizes and powers for testing τ in Design IV when $\mathbf{e} \neq \mathbf{0}$ and $n = 20$

ρ	c_1	c_2	Procedure	type	$\alpha = .01$		$\alpha = .05$		$\alpha = .1$	
					size	power	size	power	size	power
.3	1	1	prx	sys	0.013	0.285	0.048	0.533	0.095	0.673
				uns	0.006	0.178	0.041	0.471	0.092	0.616
			inc	sys	0.013	0.285	0.048	0.533	0.095	0.673
				uns	0.007	0.234	0.044	0.499	0.095	0.649
	1	4	prx	sys	0.009	0.089	0.037	0.255	**0.073**	0.358
				uns	0.011	0.105	0.049	0.267	0.103	0.404
			inc	sys	0.009	0.089	0.037	0.255	**0.073**	0.358
				uns	0.014	0.122	0.055	0.291	0.109	0.423
	4	1	prx	sys	**0.002**	0.071	**0.022**	0.245	**0.054**	0.375
				uns	0.005	0.084	0.039	0.308	0.092	0.446
			inc	sys	**0.002**	0.071	**0.022**	0.245	**0.054**	0.375
				uns	0.005	0.111	0.047	0.337	0.098	0.477
.7	1	1	prx	sys	0.013	0.672	0.056	0.864	0.099	0.928
				uns	0.010	0.538	0.043	0.814	0.090	0.907
			inc	sys	0.013	0.672	0.056	0.864	0.099	0.928
				uns	0.013	0.600	0.048	0.838	0.095	0.916
	1	4	prx	sys	**0.003**	0.209	**0.029**	0.454	**0.067**	0.601
				uns	0.005	0.248	0.042	0.535	0.089	0.681
			inc	sys	**0.003**	0.209	**0.029**	0.454	**0.067**	0.601
				uns	0.008	0.279	0.045	0.554	0.092	0.692
	4	1	prx	sys	0.007	0.212	**0.024**	0.473	**0.052**	0.641
				uns	0.009	0.284	0.054	0.570	0.105	0.745
			inc	sys	0.007	0.212	**0.024**	0.473	**0.052**	0.641
				uns	0.011	0.326	0.056	0.598	0.110	0.760

Note:

prx—Approach utilizing proxy data

inc—Incomplete data procedure (Carriere, 1994a and 1999)

sys—Compound symmetry pattern of covariance matrix

uns—Unstructured covariance pattern

Table 4.10: Empirical sizes and powers for testing γ in Design IV when $\mathbf{e} \neq \mathbf{0}$ and $n = 20$

ρ	c_1	c_2	Procedure	type	$\alpha = .01$		$\alpha = .05$		$\alpha = .1$	
					size	power	size	power	size	power
.3	1	1	prx	sys	0.010	0.200	0.051	0.410	0.107	0.544
				uns	0.005	0.141	0.050	0.366	0.103	0.509
			inc	sys	0.010	0.200	0.051	0.410	0.107	0.544
				uns	0.006	0.173	0.058	0.397	0.110	0.529
	1	4	prx	sys	**0.028**	0.108	**0.093**	0.252	**0.163**	0.352
				uns	0.009	0.041	0.046	0.162	0.108	0.261
			inc	sys	**0.028**	0.108	**0.093**	0.252	**0.163**	0.352
				uns	0.010	0.044	0.050	0.158	0.101	0.272
	4	1	prx	sys	**0.021**	0.106	**0.082**	0.238	**0.146**	0.344
				uns	0.005	0.046	0.042	0.160	0.098	0.269
			inc	sys	**0.021**	0.106	**0.082**	0.238	**0.146**	0.344
				uns	0.010	0.063	0.049	0.183	0.110	0.282
.7	1	1	prx	sys	0.015	0.472	0.054	0.741	0.094	0.827
				uns	0.012	0.369	0.042	0.673	0.094	0.792
			inc	sys	0.015	0.472	0.054	0.741	0.094	0.827
				uns	0.014	0.420	0.042	0.700	0.095	0.812
	1	4	prx	sys	**0.033**	0.211	**0.093**	0.388	**0.174**	0.493
				uns	0.009	0.081	0.049	0.272	0.093	0.391
			inc	sys	**0.033**	0.211	**0.093**	0.388	**0.174**	0.493
				uns	0.009	0.091	0.050	0.279	0.087	0.398
	4	1	prx	sys	**0.028**	0.152	**0.076**	0.317	**0.134**	0.438
				uns	0.007	0.065	0.042	0.220	0.088	0.346
			inc	sys	**0.028**	0.152	**0.076**	0.317	**0.134**	0.438
				uns	0.015	0.088	0.048	0.249	0.095	0.371

Note:
prx—Approach utilizing proxy data
inc—Incomplete data procedure (Carriere, 1994a and 1999)
sys—Compound symmetry pattern of covariance matrix
uns—Unstructured covariance pattern

Table 4.11: Proportions of proxy data that generate estimators for τ with $c \times 100\%$ efficiency in Design II

c	r_2	r_1					
		$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho \to 1.0$
.8	.50	.25	.22	.14			
	.40	.40	.38	.33	.23	.07	
	.30	.47	.46	.42	.35	.24	.06
	.20	.52	.51	.48	.42	.33	.20
	.10	.55	.54	.51	.46	.38	.28
	.00	.57	.56	.53	.49	.42	.33
.9	.30	.06	.05	.02			
	.20	.20	.19	.16	.12	.06	
	.10	.28	.27	.25	.21	.16	.10
	.00	.33	.32	.31	.28	.23	.18

Note:
r_1 is the proportion of proxy information in the second period of the sequences AB and BA.
r_2 is the proportion of proxy information in the second period of the sequences AA and BB.
The entries below the horizontal lines in each row of c correspond to the case $r_2 < r_1$.

Table 4.12: Proportions of proxy data that generate estimators for τ with $c \times 100\%$ efficiency in Design IV

c	r_1					
	$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho = 1.0$
.8	.46	.48	.48	.49	.49	.50
.9	.26	.27	.29	.29	.30	.30

Note:
r_1 is the proportion of proxy information in the third period of the sequences ABB and BAA.

Table 4.13: Proportions of proxy data that generate estimators for τ with $c \times 100\%$ efficiency in Design VI

c	r_2	r_1					
		$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho = 1.0$
.8	.50	.68	.43	.23	.05		
	.40	.75	.54	.38	.25	.15	.06
	.30	.80	.63	.49	.39	.31	.25
	.20	.85	.70	.58	.50	.43	.38
	.10	.89	.75	.65	.58	.52	.48
	.00	.92	.80	.71	.64	.59	.56
.9	.20	.38	.28	.19	.12	.07	.03
	.10	.46	.38	.31	.27	.23	.20
	.00	.52	.46	.41	.38	.35	.33

Note:

r_1 is the proportion of proxy information in the third period of the sequences ABB and BAA.

r_2 is the proportion of proxy information in the third period of the sequences AAB and BAA.

Chapter 5

Multiple Imputation for Repeated Measures Data

5.1 Introduction

In Chapter 4, we showed that using proxy information for missing data has some advantages over incomplete data analysis methods, especially when a large proportion of data is missing. As discussed in Chapter 4, the use of proxy information can be regarded as a more informative single imputation method than the ones commonly used. Comparisons between multiple imputation and single imputation strategies (Rubin, 1979, 1987) have demonstrated that single imputation generally underestimates the variability of the missing data and thus produces inefficient estimators, since the single imputed value cannot represent the uncertainty about the missing data. We can easily obtain multiply-imputed data using imputation strategies; however, it is usually not feasible to obtain multiple proxy information. To further but indirectly study the merits of using proxy information, we evaluate the efficiency of a multiple imputation strategy by comparing it with incomplete data analysis methods.

The multiple imputation strategy was proposed by Rubin in 1978. Since then, it has frequently been applied in various fields of data analysis (Rubin, 1987; Rubin and Schenker, 1986 and 1991; Little and Rubin, 1987; Taylor et al., 1990; Dorey, Little and Schenker, 1993; Heitjan and Little, 1991; Heitjan and Landis, 1994; James, 1995; etc.). The multiple imputation strategy rectifies the disadvantages of the single imputation method (Rubin, 1987). It becomes more efficient by repeating the single imputation procedure for each of the complete data sets. Valid inferences can be obtained by combining complete-data inferences, as the multiple imputations represent repeated random draws under a given model for non-response.

Other widely applied methods for dealing with incomplete data problems include the jackknife and bootstrap methods (Miller, 1974; Efron and Tibshirani, 1993; Efron, 1994), data augmentation (Tanner and Wong, 1987) and the Gibbs sampler (e.g., Gelfand and Smith, 1990; Gelman and Rubin, 1992). The multiple imputation method, like the common methods listed above, is based on the use of simulation. However, with multiple imputation, simulation is used only to handle the missing information, not the observed data. This difference provides a distinct advantage over other methods.

Like most statistical analysis methods, the multiple imputation method is not valid for all cases (Fay, 1991, 1992). In addition, the multiple imputation method has disadvantages such as the need for high-level computation and large memory space for storing the multiply-imputed data.

Further, the performance of this method in small-sample repeated measures data has not been studied (Carriere, 1997). Richardson and Flack (1996) proposed multiple imputation strategies for two-treatment three-period crossover design data, and compared it with the single imputation method and the standard analysis using only complete subset data. The sample size they used for simulation was relatively large—at least 20 subjects in each sequence.

In this chapter, we develop a multiple imputation strategy for small-sample repeated measures data in the presence of treatment effects. This chapter is organized as follows. Section 2 briefly introduces the multiple imputation theory. In Section 3, we propose a multiple imputation strategy for small-sample repeated measures data. Section 4 focuses on a comparison between the incomplete data analysis approach (Carriere, 1994a and 1999) and the proposed multiple imputation approach, with respect to the size and power of their testing procedures. Section 5 provides concluding remarks.

5.2 Multiple imputation theory

Carriere (1997) provides a comprehensive review of multiple imputation based on Rubin (1987). Multiple imputation involves drawing the missing values from the posterior distribution $f(\mathbf{y}_{mis}|\mathbf{y}_{obs})$, where $\mathbf{y}_{mis}$ is the missing portion of the data and $\mathbf{y}_{obs}$ is the observed data. Then the posterior of the parameter of interest $\boldsymbol{\theta}$, a $q \times 1$ vector, can be obtained by averaging the posterior of the complete data over the predictive distribution of the missing data, i.e., $\int g(\boldsymbol{\theta}|\mathbf{y}_{obs}, \mathbf{y}_{mis}) f(\mathbf{y}_{mis}|\mathbf{y}_{obs}) d\mathbf{y}_{mis}$. For each draw of the data, a complete data set $\mathbf{y}^{(i)} = (\mathbf{y}_{obs}, \mathbf{y}_{mis}^{(i)})$ is obtained by combining the observed data set $\mathbf{y}_{obs}$ and the imputed data set for the missing values $\mathbf{y}_{mis}^{(i)}$, $i = 1, \ldots, M$. Standard inference methods can be applied to each of the complete data sets $\mathbf{y}^{(i)}$. Then the M complete-data set analyses are combined to give a repeated-imputation inference.

Let $\hat{\boldsymbol{\theta}}_{(i)}$ and $\mathbf{U}_{(i)}, i = 1, \ldots, M$ be the estimators and their associated variances for the parameter of interest $\boldsymbol{\theta}$ based on M complete data sets. The estimator of $\boldsymbol{\theta}$ based on multiply-imputed data is

$$\overline{\boldsymbol{\theta}}_M = \sum_{i=1}^{M} \hat{\boldsymbol{\theta}}_{(i)}/M \tag{5.2.1}$$

and its variance is estimated as

$$V(\overline{\boldsymbol{\theta}}_M) = \mathbf{T}_M = \overline{\mathbf{U}}_M + (1 + M^{-1})\mathbf{B}_M, \tag{5.2.2}$$

which is composed of two elements (Rubin, 1987):

(a) the within-imputation variance

$$\overline{\mathbf{U}}_M = \sum_{i=1}^{M} \mathbf{U}_{(i)}/M \tag{5.2.3}$$

and (b) the between-imputation variance

$$\mathbf{B}_M = \sum_{i=1}^{M} (\hat{\boldsymbol{\theta}}_{(i)} - \overline{\boldsymbol{\theta}}_M)(\hat{\boldsymbol{\theta}}_{(i)} - \overline{\boldsymbol{\theta}}_M)^T/(M-1). \tag{5.2.4}$$

Consider a linear transformation $\eta = \boldsymbol{l}^T\boldsymbol{\theta}$ by a vector $\boldsymbol{l}$. Then, the approximate distribution for η is obtained via

$$(\eta - \overline{\eta}_M)[\boldsymbol{l}^T\mathbf{T}_M\boldsymbol{l}]^{-1/2} \sim t_v, \tag{5.2.5}$$

where $\overline{\eta}_M = \boldsymbol{l}^T\overline{\boldsymbol{\theta}}_M$ and the degree freedom v is

$$v = (M-1)r_M^{-2} \tag{5.2.6}$$

based on a Satterthwaite approximation (Rubin and Schenker 1986; Rubin

1987). The ratio $r_M = (1 + M^{-1})tr(\mathbf{B}_M\mathbf{T}_M^{-1})/q$ estimates the fraction of information on $\boldsymbol{\theta}$ that is missing due to non-response, which is no larger than the fraction of all missing data. The variance form with multiple imputation in Rubin (1979) does not include the factor $(1 + M)/M$. This adjustment is an improvement for modest M. It has been shown by substantial empirical Work (for example, Rubin 1979, 1998) that multiple imputation with $M = 3$ or 5 works well with typical fractions of missing data in surveys (<30%). For more extensive results on p values, see Li, Raghunathan and Rubin (1991); Li et al. (1991); and Meng and Rubin (1992).

The inference outlined above is based on a Bayesian framework that assumes that imputed values are independent draws from the posterior distribution of the missing values. Further, the size of the complete subset data is assumed to be large, so that it is effective to set the degrees of freedom for the test statistics at infinity. When the sample size is small, setting the degrees of freedom at infinity is no longer satisfactory, especially when the fractions of missing data are large. Barnard and Rubin (1999) provide a principle adjustment to the degrees of freedom v in (5.2.6), such that the resulting degrees of freedom is always less than the degrees of freedom in the complete data set. The adjusted degrees of freedom is

$$\tilde{v} = v_0\{[f(v_0)(1 - r_M)]^{-1} + \frac{v_0}{v}\}^{-1}, \tag{5.2.7}$$

where $f(v_0) = (v_0 + 1)/(v_0 + 3)$, and v_0 is the degrees of freedom based on the complete subset data.

Rubin (1987) claimed that the inference based on multiply-imputed data sets is valid if the multiple imputation strategy is proper, under the definition given in Rubin (1987). However, improper multiple imputations can still lead to valid repeated-imputation inferences (Rubin and Schenker, 1987). Theoretical work (Meng, 1994) and substantial empirical work (e.g., Rubin and Schenker, 1987; Schenker, Treiman and Weidman, 1993) support the claim that repeated-imputation inferences are confidence-valid, even if some important predictors are left out of the multiple imputation strategy—if the fraction of missing values is not too large. This conclusion provides grounds for using simpler, but possibly improper, multiple imputation strategies.

5.3 Multiple imputation for repeated measures data

Consider model (4.2.1) with $\mathbf{e} = \mathbf{0}$ and $\boldsymbol{\nu} = \mathbf{0}$. Richardson and Flack (1996) proposed a multiple imputation strategy for the data from a two-treatment three-period design. The idea is to obtain regression models for the data from periods two and three based on data from the previous periods and apply the multiple imputation strategy as illustrated in Example 5.1 (Rubin, 1987) to each of the regression models. Specifically, Richardson and Flack (1996) imputed the missing data as follows for the three-period design.

Step 1: Obtain the regression model for the second and third periods as

$$\begin{aligned} y_{ijk} &= \mathbf{x}_{ik}\boldsymbol{\beta} + (\mathbf{y}_{jk(i-1)} - \mathbf{X}_{k(i-1)}\boldsymbol{\beta})^T\boldsymbol{\beta}_{i1} + \varepsilon_{ijk} \\ &= \beta_{ik0} + \mathbf{y}^T_{jk(i-1)}\boldsymbol{\beta}_{i1} + \varepsilon_{ijk} \\ &= \mathbf{z}_{jk}\boldsymbol{\beta}_{ik} + \varepsilon_{ijk} \qquad (5.3.1) \end{aligned}$$

for $i = 2, 3$ and $k = 1, 2$, where $\mathbf{y}_{jk(2)} = (y_{1jk}, y_{2jk})^T$ and $\mathbf{y}_{jk(1)} = y_{1jk}$, $\mathbf{z}_{jk} = (1, \mathbf{y}^T_{jk(i-1)})$ and $\boldsymbol{\beta}_{ik} = (\beta_{ik0}, \boldsymbol{\beta}^T_{i1})^T$ and get the usual least-square estimator $\hat{\boldsymbol{\beta}}_{ik}$. Here, $\mathbf{x}_{ik}$ is the i^{th} row of the design matrix $\mathbf{X}_k$ and $\mathbf{X}_{k(i-1)}$ is the submatrix of $\mathbf{X}$ including the first $i-1$ rows, $\boldsymbol{\beta} = (\mu, \pi_1, \pi_2, \tau, \gamma, \lambda)^T$. Let SSE_i denote the sum of squared error from the regression model for the response in the i^{th} period and let $V_{ik} = [\sum_j \mathbf{z}_{jk}\mathbf{z}^T_{jk}]^{-1}$.

Step 2: Draw a chi-square random variable g_i with $N^{(i)} - d_i$ degrees of freedom, where d_i is the number of elements in $\boldsymbol{\beta}_{ik}$ and $N^{(i)}$ is the number of subjects completing the first i periods. Then define $\sigma_i^{*2} = SSE_i/g_i$.

Step 3: Draw d_i independent $N(0, 1)$ variates to create a d_i- component vector r_i, and update $\hat{\boldsymbol{\beta}}_{ik}$ as $\boldsymbol{\beta}^*_{ik} = \hat{\boldsymbol{\beta}}_{ik} + \sigma_i^* {V_{ik}}^{1/2} r_i$.

Step 4: Impute the missing data y_{ijk} by $\mathbf{z}_{jk}\boldsymbol{\beta}^*_{ik} + r^*_0$, where r^*_0 is randomly selected from residuals $\{y_{ijk} - \mathbf{z}_{jk}\boldsymbol{\beta}^*_{ik}\}$ with replacement and with equal probability, and j runs over all subjects completing the first i periods in sequence k .

Richardson and Flack (1996) obtained multiply-imputed data sets by repeating Steps 1-4 several times. They showed that their strategy works well for large sample data in simulation. However, their work is limited in the scope of sample sizes and missing proportions. Further, their regression approach can fail to impute data in the last several periods, since the degrees of freedom to generate the chi-square random variable can be 0.

Assuming a monotonic missing pattern, we propose an imputation strategy that would be simple and also work well for small-sample data, improving the shortcomings noted above. Rather than relying on the regression model, we simply impute the missing values using the conditional distribution of the missing data, given the observed data in previous periods. Consequently, the proposed multiple imputation strategy is robust against model assumptions.

For the complete subset data, we assume that $\mathbf{y}_{(p_1)jk}$ is distributed as multivariate normal with mean $\boldsymbol{\mu}_{(p_1)k}$ and covariance matrix $\boldsymbol{\Sigma}_{11}$, using the notation established in Chapter 4. Thus, the conditional distribution of $y_{p_1+1,jk}$ given $\mathbf{y}_{(p_1)jk}$ is normal with mean $\mu_{p_1+1,k} + \boldsymbol{\sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_{(p_1)jk} - \boldsymbol{\mu}_{(p_1)k})$, and variance

$\sigma_{22} - \boldsymbol{\sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\sigma}_{12}$, where $\boldsymbol{\Sigma}_{11}$ is the submatrix of $\boldsymbol{\Sigma}$ with the first p_1 rows and p_1 columns, $\boldsymbol{\sigma}_{21}$ is that with the $(p_1+1)^{th}$ row and the first p_1 columns, with $\boldsymbol{\sigma}_{12} = \boldsymbol{\sigma}_{21}^T$, and σ_{22} is the $(p_1+1)^{th}$ diagonal element of $\boldsymbol{\Sigma}$. For the subsequent periods, we applied the above steps by replacing p_1 with $p_1 + 1$.

The following six steps made up our imputation strategy. We adopted and extended the imputation strategy for a univariate normal model as described in Example 3.2 (Rubin, 1987) in order to apply it to a multivariate normal model.

Step 1: With the usual least-squares estimators for the mean and the covariance matrices, we have that $y_{p_1+1,jk}$ given $\mathbf{y}_{(p_1)jk}$ is approximately distributed as normal with mean $\hat{\mu}_{p_1+1,k} = \overline{y}^{(l)}_{p_1+1.k} + \mathbf{s}_{21}\mathbf{S}_{11}^{-1}(\mathbf{y}_{(p_1)jk} - \overline{\mathbf{y}}^{(1)}_{(p_1).k})$ (see Chapter 4 for the notation), $k = 1, \ldots, s$, and variance $\hat{\sigma}^2 = s_{22} - \mathbf{s}_{21}\mathbf{S}_{11}^{-1}\mathbf{s}_{12}$, where l is the index for the level the $(p_1+1)^{th}$ period belongs to, that is, $R_{l-1} < p_1 + 1 \leq R_l$, $\mathbf{S}$ is the sample variance using the complete data subset, and $\mathbf{S}_{11}$, $\mathbf{s}_{12}$, $\mathbf{s}_{21}$ and s_{22} are the submatrices defined in the same way as $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\sigma}_{12}$, $\boldsymbol{\sigma}_{21}$ and σ_{22}, respectively.

Step 2: Draw a chi-square random variable g with degrees of freedom $N^{(l)} - s$ and let $\sigma^* = \hat{\sigma}(N^{(l)} - s)/g$.

Step 3: Draw a random variable from a standard normal distribution, say z, and let $\mu^{*}_{p_1+1,k} = \hat{\mu}_{p_1+1,k} + \sigma^{*} z / \sqrt{N_k^{(l)}}$.

Step 4: Draw a random variable z from a standard normal distribution, and impute for the missing values in the period of $p_1 + 1$ as $y_{p_1+1,jk} = \mu^{*}_{p_1+1,k} + \sigma^{*} z$. Repeat Step 4 for all missing components in the period $p_1 + 1$.

Step 5: Treat the imputed values as if they are observed. Repeat Steps 1-4 for the next periods with missing values, with p_1 replaced by $p_1 + 1$.

Step 6: Repeat Steps 1-5 M times for all sequences to create M multiply-imputed data sets.

The proposed imputation strategy implies that the imputed values in previous periods are treated as if they were observed when they are used to impute missing values in the current period. The idea is similar to the sequential imputation method (Kong, Liu and Wong, 1994). The rationale lies in that repeated measures data are generally obtained at different times of measurement. Thus, data in different time periods can be regarded as if they became available sequentially. However, each of the imputed "complete" data sets contributes to the repeated-imputation inference equally, which is not the case in sequential imputation.

Applying the multiple inference method outlined in Section 2, we obtained the estimator $\hat{\boldsymbol{\beta}}_{MI}$, using (5.2.1), and the covariance matrix $\mathbf{V}_{MI}$, using (5.2.2). In the next section, we demonstrate the performance of the proposed multiple imputation strategy by comparing the size and power of the corresponding testing procedure with those of the incomplete data analyses of Carriere (1994a and 1999).

5.4 Simulations

As stated in Chapter 4, the hypotheses of interest are those of no treatment effects $H_{01} : \tau = 0$, and no residual effects $H_{02} : \gamma = 0$. The test statistic we considered is

$$t_\theta = \frac{\hat{\theta}}{v\hat{a}r(\hat{\theta})^{1/2}},$$

where $v\hat{a}r(\hat{\theta})$ is the estimated variance for estimation $\hat{\theta}$ using multiply-imputed data and $\hat{\theta}$ is either $\hat{\tau} = \boldsymbol{l}_1^T \hat{\boldsymbol{\beta}}_{MI}$ or $\hat{\gamma} = \boldsymbol{l}_2^T \hat{\boldsymbol{\beta}}_{MI}$, with $\boldsymbol{l}_i$ being a vector of 0 and 1, $i = 1, 2$. Under the assumption of normality, the statistic t_θ is expected to have an approximate t-distribution with degrees of freedom $\tilde{v}$ as in (5.2.7), based on the multiply-imputed data.

Carriere (1999) proposed a t test for treatment effects and residual effects using the estimator, as summarized in Section 3, Chapter 4. The degrees of freedom Carriere (1999) proposed is $(p-1)(N^{(L)} - s)$ for both τ and γ under the assumption that the covariance matrix is compound symmetric, while it is $N^{(L)} - s$ for treatment effects and $(N + N^{(2)} - 2s - p_1p_2)/2$ for residual effects under an unspecified covariance structure assumption.

To examine the performance of these two approaches for small-sample data, we generated 1000 small samples from the model (4.2.1) using Design I (design with sequences AB and BA, as defined in Chapter 4) and Design IV (design with sequences ABB and BAA, as defined in Chapter 2) with five subjects in each sequence. We consider a case where 20% of the data is missing, assuming that the missing pattern is monotonic. In each sequence, we randomly choose one subject and delete the observations in their last period to generate missing data. We set $M = 5$ for the multiple imputation procedure. The covariance matrix $\mathbf{\Sigma}$ in model (4.2.1) is considered the same as in Carriere (1999)

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & \rho\sqrt{c_1} & \rho\sqrt{c_2} \\ & c_1 & \rho\sqrt{c_1c_2} \\ & & c_2 \end{bmatrix}$$

with two different values of within-subject correlation ($\rho = .3$ and $\rho = .7$) and two different values, 1 or 4 for c_1 and c_2. For the two-period design, the last column and row are deleted from $\mathbf{\Sigma}$. The sizes for testing H_{01} and H_{02}, and powers of the test for $H_{11} : \tau = .5$ and $H_{12} : \gamma = .5$ are calculated using both approaches. For Design IV, we calculated the powers for $H_{11} : \tau = .25$ and

$H_{12} : \gamma = .25$.

Overall, the asymptotic distribution suggested earlier appears to fit reasonably well in both approaches. Empirical sizes that do not agree with the nominal levels are highlighted in Tables 5.1-5.3. The 95% confidence intervals for these empirical levels are (0.004, 0.016), (0.036, 0.064) and (0.081, 0.119) for nominal level .01, .05 and .10, respectively.

Although the empirical sizes were satisfactory in keeping the nominal levels, we decided to adopt the α-adjusted power as a criteria for evaluating the performance of the two approaches (Kim, 1992; Cristensen and Rencher, 1997). The α-adjusted power is determined by finding the proportion of times that the p-value for a test when H_{01} or H_{02} is false is less than the α-adjusted critical level. The α-adjusted critical level is defined as the α^{th} percentile of the 1000 p-values when the null hypothesis is true. The α-adjusted power allows us to compare the power upon adjusting the empirical sizes to be the same, thereby eliminating the power advantage of procedures with higher empirical size.

Table 5.1 demonstrates the empirical sizes for testing H_{01} and H_{02}, and α-adjusted powers for testing H_{11} and H_{12}, for both the multiple imputation approach and the incomplete data analysis method for the two-period design. The performance of the multiple imputation method was similar to that of the incomplete data method (Carriere, 1994a and 1999) for all cases we considered, except when ρ is large for testing treatment effects, where the incomplete data method is more powerful.

Tables 5.2 and 5.3 show similar results for the three-period design for γ and τ. The two approaches are rather similar, although the incomplete data method is more powerful when ρ is large.

5.5 Conclusion

This chapter investigated the merits of a multiple imputation strategy over the incomplete data analytic method. We first suggested a small-sample distribution of the estimators based on multiply-imputed data, assuming that the data are distributed as a multivariate normal. The strategy we proposed does not rely on the regression model, which makes it robust against the violation of the assumed model. The imputation strategy is simple and easy to implement. Once the missing data are filled in, standard statistical inferences can be drawn, based on the multiply-imputed complete data sets. No special software is required.

Next, we used simulation studies to show that the multiple imputation method performs reasonably well in most of the cases we considered. However, we generally did not find it superior to the alternative method (Carriere, 1999).

In the previous chapter, we concluded that using proxy information can be a competitive alternative to missing data analysis. This conclusion leads us to investigate multiple imputation to see if this approach would provide any further improvement. However, there does not appear to be much benefit in using the multiple imputation method, considering the time and effort required. This investigation finds that the multiple imputation method is not superior to the missing data analysis method utilizing all available data. Considering the advantages of utilizing proxy information for missing values, as discussed in Chapter 4, we conclude that one should try to collect proxy data, which is a more personal and informative method of imputing for missing data than using conventional single or multiple imputations.

Table 5.1: Empirical sizes and α-adjusted powers for testing τ and γ in Design I.

	ρ	c_1	Method	Type	$\alpha = .01$		$\alpha = .05$		$\alpha = .1$	
					size	power	size	power	size	power
γ	.3	1	INC	sys	0.011	0.025	0.051	0.121	0.104	0.229
				uns	0.013	0.028	0.051	0.135	0.114	0.223
			MI	sys	0.007	0.032	0.037	0.130	0.084	0.229
				uns	0.013	0.021	0.044	0.124	0.109	0.204
		4	INC	sys	0.014	0.009	0.054	0.082	0.114	0.147
				uns	0.016	0.010	0.057	0.066	0.112	0.148
			MI	sys	0.013	0.010	0.060	0.058	0.109	0.152
				uns	0.013	0.007	0.051	0.075	0.104	0.146
	.7	1	INC	sys	0.011	0.017	0.047	0.106	0.094	0.199
				uns	0.008	0.023	0.042	0.116	0.092	0.189
			MI	sys	0.007	0.019	0.036	0.112	0.089	0.181
				uns	0.007	0.026	0.038	0.116	0.090	0.188
		4	INC	sys	0.011	0.016	0.051	0.067	0.112	0.125
				uns	0.006	0.021	0.048	0.074	0.100	0.131
			MI	sys	0.008	0.023	0.048	0.080	0.110	0.124
				uns	0.004	0.027	0.036	0.083	0.086	0.144
τ	.3	1	INC	sys	0.011	0.182	0.059	0.509	0.102	0.691
				uns	0.010	0.189	0.056	0.497	0.099	0.701
			MI	sys	0.008	0.201	0.045	0.511	0.091	0.693
				ubs	0.010	0.194	0.050	0.498	0.094	0.676
		4	INC	sys	0.014	0.078	0.061	0.235	0.116	0.367
				uns	0.011	0.075	0.060	0.232	0.107	0.351
			MI	sys	0.010	0.082	0.054	0.246	0.116	0.352
				uns	0.007	0.098	0.054	0.211	0.115	0.332
	.7	1	INC	sys	0.008	0.558	0.050	0.871	0.094	0.951
				uns	0.010	0.541	0.051	0.865	0.092	0.946
			MI	sys	0.006	0.506	0.047	0.831	0.087	0.927
				uns	0.011	0.470	0.045	0.826	0.089	0.923
		4	INC	sys	0.004	0.185	0.061	0.333	0.115	0.505
				uns	0.004	0.170	0.058	0.348	0.111	0.490
			MI	sys	0.005	0.139	0.055	0.342	0.105	0.499
				uns	0.005	0.167	0.050	0.326	0.097	0.485

MI—Multiple imputation approach
INC—Incomplete data procedure (Carriere, 1994a and 1999)

Table 5.2: Empirical sizes and α-adjusted powers for testing γ in Design IV.

ρ	c_1	Methods	Type	$\alpha = .01$		$\alpha = .05$		$\alpha = .1$	
				size	power	size	power	size	power
.3	1	INC	sys	0.011	0.063	0.056	0.182	0.115	0.283
	1		uns	0.006	0.053	0.050	0.171	0.109	0.272
		MI	sys	0.005	0.089	0.037	0.195	0.081	0.333
			uns	0.007	0.054	0.039	0.189	0.083	0.299
	1	INC	sys	**0.025**	0.032	**0.105**	0.088	**0.179**	0.156
	4		uns	0.010	0.029	0.060	0.085	0.112	0.155
		MI	sys	**0.029**	0.024	**0.111**	0.090	**0.176**	0.151
			uns	0.012	0.022	0.057	0.093	0.115	0.165
	4	INC	sys	**0.020**	0.012	**0.076**	0.112	**0.137**	0.193
	1		uns	0.011	0.011	0.050	0.090	0.097	0.178
		MI	sys	0.012	0.021	0.053	0.109	0.093	0.224
			uns	0.007	0.017	**0.033**	0.112	**0.076**	0.201
.7	1	INC	sys	0.010	0.175	0.051	0.428	0.105	0.555
	1		uns	0.011	0.093	0.048	0.387	0.097	0.497
		MI	sys	0.007	0.201	0.043	0.413	0.082	0.574
			uns	0.007	0.141	0.037	0.366	0.082	0.523
	1	INC	sys	**0.025**	0.054	**0.086**	0.146	**0.151**	0.237
	4		uns	0.005	0.055	0.041	0.151	0.091	0.250
		MI	sys	**0.029**	0.051	**0.093**	0.149	**0.163**	0.232
			uns	0.004	0.044	0.052	0.130	0.100	0.224
	4	INC	sys	**0.023**	0.036	**0.079**	0.151	**0.149**	0.253
	1		uns	0.010	0.026	0.045	0.139	0.094	0.233
		MI	sys	0.004	0.074	**0.028**	0.200	**0.067**	0.322
			uns	0.005	0.040	**0.021**	0.181	**0.052**	0.294

MI—Multiple imputation approach
INC—Incomplete data procedure (Carriere, 1994a and 1999)

Table 5.3: Empirical sizes and α-adjusted powers for testing τ in Design IV.

ρ	c_1	Procedure	Type	$\alpha = .01$		$\alpha = .05$		$\alpha = .1$	
				size	power	size	power	size	power
.3	1	INC	sys	0.011	0.106	0.059	0.257	0.106	0.408
	1		uns	0.006	0.091	0.050	0.245	0.116	0.331
		MI	sys	0.007	0.100	0.057	0.255	0.111	0.383
			uns	0.004	0.100	0.051	0.232	0.100	0.372
	1	INC	sys	0.007	0.059	0.043	0.178	0.088	0.274
	4		uns	0.009	0.059	0.051	0.173	0.104	0.277
		MI	sys	0.007	0.056	0.045	0.169	0.091	0.269
			uns	0.009	0.054	0.049	0.191	0.106	0.281
	4	INC	sys	**0.003**	0.067	**0.029**	0.183	**0.063**	0.297
	1		uns	0.010	0.051	0.051	0.152	0.106	0.263
		MI	sys	**0.002**	0.065	**0.026**	0.156	**0.064**	0.275
			uns	0.009	0.065	0.051	0.153	0.110	0.263
.7	1	INC	sys	0.004	0.323	0.039	0.585	0.090	0.717
	1		uns	0.008	0.215	0.045	0.499	0.088	0.692
		MI	sys	0.009	0.279	0.036	0.554	0.086	0.694
			uns	0.005	0.235	0.042	0.516	0.083	0.663
	1	INC	sys	0.008	0.105	0.036	0.280	**0.072**	0.431
	4		uns	0.008	0.076	0.045	0.293	0.094	0.415
		MI	sys	0.007	0.099	0.043	0.257	0.082	0.406
			uns	0.007	0.088	0.053	0.253	0.111	0.371
	4	INC	sys	**0.003**	0.110	**0.026**	0.326	**0.055**	0.458
	1		uns	0.005	0.104	0.045	0.286	0.082	0.455
		MI	sys	**0.002**	0.120	**0.022**	0.304	**0.053**	0.432
			uns	0.007	0.106	0.055	0.259	0.104	0.403

MI—Multiple imputation approach

INC—Incomplete data procedure (Carriere, 1994a and 1999)

Chapter 6

Concluding Remarks

Repeated measurements designs are often employed to measure the efficacy of several treatments or interventions and to make appropriate recommendations about the "best" treatment or intervention. Statistical analyses based on the data from these designs are carried out to guide decision-making processes.

In these empirical investigations, many statistical issues arise. This thesis is a study of design and analysis issues in typical experiments that use repeated measurement designs, especially issues pertaining to data based on small samples.

6.1 Main contributions

1. **Construction of optimal designs under a general model with ran-**

dom subject effects and autoregressive error structure. Many investigators have constructed optimal designs under various models. However, these models are simple and the designs constructed cannot be generalized to other situations. We considered a more general model that accommodates serially correlated measurement error and random subject effects. All other models are special cases of this general model. Optimal designs are highly dependent on the autoregressive correlation coefficient ϕ and the ratio of within- and between-subject variances for the measurement error, ρ. However, for two-treatment two-period designs, we found that a design with an equal number of subjects in the sequences AA, AB and their duals is the "best" design for estimating τ and γ.

For two-treatment three-period designs, we found that the design with an equal allocation of subjects to each of the sequences ABB, BAA, AAB and BBA is nearly as efficient for treatment effects as the optimal designs across the range of ρ and ϕ. For residual effects, a design with an equal allocation of subjects to the sequences ABB and BAA generally performs well, except for extremely negative ϕ and small ρ. The four-sequence design has been suggested as the best compromised design under various assumptions about residual effects (Carriere, 1994). The two-sequence design is the universally optimal design under an equicorrelated covariance structure with first-order residual effects (Laska et al., 1983; Kershner, 1986; Carriere, 1994).

When one more period is added into the experiment, a design with an equal number of subjects assigned to AABB, ABBA and their duals, BBAA and BAAB, is found to be the most robust design for both treatment effects and residual effects. This design has also been recommended as the "best" design under an autoregressive error model with fixed subject effects (Matthews, 1987).

2. **Construction of designs adaptive to data.** The study subjects are naturally heterogeneous and the associated covariance matrices are typically unknown. Further, the subjects enter the experiment in sequence. No known optimal design exists under this situation. To construct a design suitable for these situations, we first entered a few subjects in the experiment and obtained the information matrix for treatment effects for t-treatment p-period designs. By using and updating the information matrix repeatedly, allocation rules of subjects to sequences were derived in such a way that the loss function of the additional information from the new set of subjects is minimized, based on a criterion. Compared to the approach that ignores heterogeneity of subjects, this adaptive approach is up to 33% more efficient. However, as expected, such consideration

of heterogeneity requires a considerably large number of subjects to be included in the experiment. Otherwise, the distinct covariance matrices are not distinguishable. In general, adaptive designs are more intuitive and realistic, and no doubt more efficient than designs constructed on the basis of unrealistic assumptions.

3. **Using proxy information for missing data.** We built a model that incorporated possible bias and heterogeneity in responses when some data are provided by proxies. We then developed an inferential procedure with small-sample repeated measures data involving proxy information. As expected, when the proxy and the observed data share the same covariance structure, the resulting testing procedure is more powerful than the incomplete data method. When heterogeneity is present, this testing procedure is not as powerful as incomplete data methods, especially when the missing proportion and thus the amount of proxy data is small, i.e., 20%. However, when the missing proportion increases to 50%, the method utilizing proxy is superior to the incomplete data method.

   In addition, general conditions were obtained under which the use of proxy is nearly as efficient as when all data are actual. When the proxy and observed data share the same covariance matrix, using over 50% proxy information can still produce estimators that are at least 90% as efficient as those obtained when all data is actual.

4. **Multiple imputation approach for the missing data problem in repeated measurement designs.** We proposed a multiple imputation strategy for the missing data problem in repeated measurement designs. This strategy performs reasonably well even when the sample size is small. The general conclusion is that, considering the work required, multiple imputation of missing data is not worthwhile in repeated measurement designs. The incomplete data method was found to be just as good as the imputation approach, without the extra work. However, as per the results from Chapter 4, whenever possible, proxy information should be used to obtain better and more powerful results.

6.2 Future direction

There are a number of issues that still need to be resolved. One of the main contributions of this thesis is the new knowledge gained about the adaptive construction of repeated measurement designs, which generate more efficient estimators for parameters of interest. However, more subjects will need to be involved in experiments before the proposed strategy becomes desirable and useful. The number of subjects required for a design depends on the level of heterogeneity among subjects. The more heterogeneous the subjects, the fewer the subjects required. In simulation studies conducted for this thesis, the required number of subjects was usually over 30.

We are working towards improving the allocation rules so that an efficient design can be generated without requiring an excessively large sample size. One possibility is to incorporate more information in the loss function by utilizing available demographic or prognostic factors.

We will also explore constructing adaptive designs under a mixture model, where the study population is composed of several populations, possibly with different means and different covariance matrices.

For the multiple imputation approach, we assumed that the missing data occurred at random. Further work is needed to examine the sensitivity of the proposed strategy to the violation of the assumed mechanism. Recently, methods based on pattern mixtures have generated much interest. We plan to extend our work to the non-randomly missing data situation. We will also compare our strategy to that using weighted estimating equations (Fitzmaurice et al., 1995; Robins et al., 1995; Rotnizky et al., 1998).

Finally, all analysis work in this thesis assumed that the data are from a multivariate normal distribution. The methods discussed might be extended to data from non-normal distributions, but first it will be necessary to explore the sensitivity of the proposed methods to violations of the model assumptions.

Bibliography

[1] Aigner, D. J. (1974) MSE dominance of least squares with errors-of-observation. *Journal of Econometrics*, 2, 365-372.

[2] Anderson, T. W. (1957) Maximum likelihood estimators for a multivariate normal distribution when some observations are missing. *Journal of American Statistical Association*, 52, 200-203.

[3] Armitage, P. (1975) *Sequential Medical Trials.* Oxford: Blackwell.

[4] Armitage, P. (1985) The search for optimality in clinical trials. *International Statistical Review*, 53, 15-36.

[5] Atkinson, A. C. (1982) Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika*, 69, 61-67.

[6] Barnard, J. O. and Rubin, D. B. (1999) Small-sample degrees of freedom with multiple imputation. *Biometrika*, 86, 4, 948-955.

[7] Barnow, B. S. (1976) The use of proxy variables when one or two independent variables are measured with error. *The American Statistician*, 30, 119-121.

[8] Bartlett, R. H., Roloff, D. W., Cornell, R. G., Andrews, A. F., Dillon, P.W., and Zwischenberger, J. B. (1985) Extracorporeal circulation in neonatal respiratory failure: a prospective randomized study. *Pediatrics* 76, 479-487.

[9] Bather, J. A. (1985) On the allocation of treatments in sequential medical trials. *International Statistical Review*, 53, 1-13.

[10] Brown, B. W. (1980) The crossover experiment for clinical trial. *Biometrics*, 36, 69-79.

[11] Carrière, K. C. (1994a) Incomplete repeated measures data analysis in the presence of treatment effects. *Journal of the American Statistical Association*, 89, 680-686.

[12] Carrière, K. C. (1994b) Crossover designs for clinical trials. *Statistics in Medicine*, 13, 1063-1069.

[13] Carrière, K. C. (1997) Comparative study on missing data analysis methods. Unpublished manuscript.

[14] Carrière, K. C. (1999) Methods for repeated measures data analysis with missing values. *Journal of Statistical Planning and Inference*, 77, 221-236.

[15] Carrière, K. C. and Reinsel, G. C. (1992) Investigation of dual-balanced crossover designs for two treatments. *Biometrics*, 48, 1157-1164.

[16] Carrière, K. C. and Reinsel, G. C. (1993) Optimal two-period repeated measurement designs with two or more treatments. *Biometrika*, 80, 924-929.

[17] Carrière, K. C. and Huang, R. (2000) Crossover designs for two-treatment clinical trials. *Journal of Statistical planning and Inference* 87, 125-134.

[18] Carrière, K. C. and Huang, R. (2001) Traditional Paradigms in Pharmacoeconomics: Consideration for Cost-Effective Designs. *In Introduction to Applied Pharmacoeconomics.* McGraw-Hill Medical Publishing Division. 19-39.

[19] Cheng, C. S. and Wu, C. F. (1980) Balanced repeated measurements designs (Corr: V11 p349). *The Annals of Statistics*, 8, 1272-1283.

[20] Christensen, W. F. and Rencher, A. C. (1997) A comparison of type I error rates an power levels for seven solutions to the multivariate Behrens-Fisher problem. *Communication in Statistics, Simulation*, 26 (4), 1251-1273.

[21] Cleophas, T.J.M. and Tavenier, P. (1994) Fundamental issues of choosing the right type of trial. *American Journal of Therapeutics*, 1, 327-332.

[22] Cochran, W.G. and Cox, G.M. (1957) *Experimental Designs*, 2dn Edition. Wiley: New York.

[23] Cook, R. J. (1995) Interim analyses in 2×2 crossover trials. *Biometrics*, 51, 932-945.

[24] Cook, R. J. (1996) Design considerations in crossover trials with a single interim analysis and serial patient entry. *Biometrics*, 52, 732-739.

[25] Dhrymes, P. J. (1978) *Introductory Econometrics.* Springer, New York.

[26] Diggle, P. J. (1988) An approach to the analysis of repeated measurements. *Biometrics*, 44, 959-971.

[27] Dorey, F. J., Little, R. J. A. and Schenker, J. L. (1993) Multiple imputation for threshold crossing data with interval censoring. *Statistics in Medicine*, 12, 1589-1603.

[28] Ebbutt, A. F. (1984) Three-period crossover designs for two treatments. *Biometrics*, 40, 219-224.

[29] Efron, B. (1971) Forcing a sequential experiment to be balanced. *Biometrika*, 58, 403-417.

[30] Efron, B. (1994) Missing data, Imputation, and the Bootstrap. *Journal of American Statistical Association*, 89, 463-478.

[31] Efron, B. and Tibsharani, R. (1993) *An introduction to the Bootstrap*. London, Chapman and Hall.

[32] Engle, B. and Walstra, P. (1991) Increasing precision or reducing expense in regression experiments by using information from a concomitant variable. *Biometrics*, 47, 13-20.

[33] Fay, R. E. (1991) A design-based perspective on missing data variance. *Proceeding of the 1991 annual research conference*, U.S. Bureau of census, 429-440.

[34] Fay, R. E. (1992) When are inferences from multiple imputation valid? ASA *Proceedings of the Section on Survey Research Methods*, 227-232.

[35] Fitzmaurice, G. M., Molenberghs, G. and Lipsitz, S. R. (1995) Regression models for longitudinal binary responses with informative drop-outs. *Journal of the Royal Statistical Society*, B, 57, 691-704.

[36] Fleiss, J. L. (1989) A critique of recent research of the two-treatment cross-over design. *Control Clinical Trials*, 20, 237-243.

[37] Flournoy, N. and Rosenberger, W. F. (1995) *Adaptive Design*. Institute of Mathematical Statistics Lecture Notes-Monograph Series, Vol. 25.

[38] Frost, P. A. (1979) Proxy variables and specification bias. *The Review of Economics and Statistics*, 61, 323-325.

[39] Gelfand, A. E. and Smith, A. F. M. (1990) Sampling-based approaches to calculating marginal densities . *Journal of the American Statistical Association*, 85, 398- 409.

[40] Gelman, A. and Rubin, D. B. (1992) Inference from iterative simulation using multiple sequences (Disc: p483-501, 503-511). *Statistical Science*, 7, 457- 472.

[41] Glynn, R. J., Laird, N. M. and Rubin, D. B. (1993) Multiple Imputation in Mixture Models for Nonignorable Nonresponse With Follow-ups, *Journal of the American Statistical Association*, Vol. 88, No. 423. (Sep., 1993), pp. 984-993.

[42] Grizzle, J. E. (1965) The two-period change-over design and its use in clinical trials (Corr: V30 p727). *Biometrics*, 21, 467- 480.

[43] Grootendorst, P., Feeny, D. and Furlong, W. (1997) Does it matter whom and how you ask? An investigation into inter- and intra-rater agreement in the 1996 Ontario Health Survey. *Journal of Clinic Epidemiology*, 50, 127-135.

[44] Hedayat, A. and Afsarinejad, K. (1975) Repeated measurements design, I. In *A Survey of Statistical Designs and Linear Models, Ed. J. N. Srivastava*, pp. 229-242. Amsterdam: North-Holland.

[45] Hedayat, A. and Afsarinejad, K. (1978) Repeated measurements design, II. *The Annals of Statistics*, 6, 619-628.

[46] Hedayat, A. and Zhao, W. (1990) Optimal two-period repeated measurements designs. *The Annals of Statistics*, 18, 1805-1816.

[47] Heitjan, D. F. and Landis, J. R. (1994) Assessing secular trends in blood pressure: A multiple-imputation approach. *Journal of the American Statistical Association*, 89, 750-759.

[48] Hogan, J. W. and Liard, N.M. (1997) Mixture models for the joint distribution of repeated measures and event time. *Statistics in Medicine*, 16, 239-258.

[49] Jalukar, V., Funk, G. F., Christensen, A. J., Karnell, L. H. and Moran, P. J. (1998) Health states following head and neck cancer treatment: patient, health-care professional, and public perspectives. *Health and Neck*, 20, 1998, 600-608.

[50] James, I. R. (1995) A note on the analysis of censored regression data by multiple imputation. *Biometrics*, 51, 358-362.

[51] Kershner, R. P. (1986) Optimal three-period two-treatment cross-over designs with and without baseline readings. *Proceedings of the Biopharmaceutical Section of the American Statistical Association*, 151-160.

[52] Kershner, Ronald P. and Federer, Walter T. (1981) Two-treatment crossover designs for estimating a variety of effects. *Journal of American Statistical Association*, 76, 612- 619.

[53] Kiefer, J. (1975) Optimal design: Variation in structure and performance under change of criterion. *Biometrika*, 62, 277- 288.

[54] Kim, S. J. (1992) A practical solution to the multivariate Behrens-Fisher problem. *Biometrika*, 79 (1), 171-176.

[55] Kong, A., Liu, J. S. and Wong, W. H. (1994) Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, 89, 278- 288.

[56] Kunert, J. (1983) Optimal design and refinement of the linear model with applications to repeated measurements designs. *The Annals Statistics*, 11, 247- 257.

[57] Kunert, J. (1984) Optimality of balanced uniform repeated measurements designs. *The Annals of Statistics*, 12, 1006- 1017.

[58] Kunert, J. (1985) Optimality repeated measurements designs for correlated observations and analysis by weighted least squares. *Biometrika*, 72, 375-389.

[59] Kunert, J. (1991) Cross-over designs for two treatments and correlated errors. *Biometrika*, 78, 315-324.

[60] Kushner, H. B. (1997) Optimality and efficiency of two-treatment repeated measurements design. *Biometrika*, 84, 455-468.

[61] Kushner, H. B. (2000) Allocations rules for adaptive repeated measurements designs. Unpublished manuscript.

[62] Laska, E., Meisner, M. and Kushner, H. B. (1983) Optimal crossover designs in the presence of carryover effects. *Biometrics*, 39, 1087-1091.

[63] Laska, E. M. and Meisner, M. (1985) A variational approach to optimal two-treatment crossover designs: Application to carryover-effect models. *Journal of the American Statistical Association*, 80, 704- 710.

[64] Li, K. H., Raghunathan, T. E. and Rubin, D. B. (1991) Large-sample significance levels from multiply imputed data using moment-based statistics and an F reference distribution. *Journal of the American Statistical Association*, 86, 1065-1073.

[65] Li, K. H., Meng, X. L., Raghunathan, T. E. and Rubin, D. B. (1991) Significance levels from repeated p-values with multiply-imputed data. *Statistica Sinica*, 1, 65-92.

[66] Little, R. J. A. and Rubin, D. B. (1987) *Statistical Analysis with Missing Data*. New York: John Wiley.

[67] Maddala, G. S. (1977) *Econometrics*. McGraw-Hill: New York.

[68] Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979) *Multivariate Analysis*. Academic Press.

[69] Matthews, J. N. S. (1987) Optimal crossover designs for the comparison of two treatments in the presence of carryover effects and autocorrelated errors. *Biometrika*, 74, 311- 320.

[70] Matthews, J. N. S. (1988) Recent developments in crossover designs. *International Statistical Review*, 56, 117- 127.

[71] Matthews, J. N. S. (1994a) Multi-period crossover trials. *Statistical Methods in Medical Research*, 3, 383-405.

[72] Matthews, J. N. S. (1994b) Modelling and optimality in the design of crossover studies for medical applications. *Journal of Statistical Planning and Inference*, 42, 89-108.

[73] McCallum, B. T. (1972) Relative asymptotic bias from errors of omission and measurement. *Econometrica*, 40, 757-758.

[74] McDonald, R. P. and Swaminathan, H. (1973) A simple matrix calculus with applications to multivariate analysis. *General Systems*, XVIII, 37-54.

[75] Meng, X. L. (1994) Multiple imputation with uncongenial sources of input. *Statistical Science*, 9, 538-574.

[76] Meng, X. L. and Rubin, D. B. (1992) Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, 79, 103-111.

[77] Miller, R. G. (1974) The Jackknife–A review. *Biometrika*, 61, 1-17.

[78] Morrison, D. F. (1970) Expectations and Variances of maximum likelihood estimates of the multivariate normal distribution parameters with missing data. *Technical Report No. 1*, Department of Statistics and Operations Research, University of Pennsylvania.

[79] Ohtani, K. (1981) On the use of a proxy variable in prediction: An MSE comparison. *The Review of Economics and Statistics*, 63, 627- 629.

[80] Parsons, S. K., Barlow, S. E., Levy, S. L., Supran, S. E. and Kaplan, S. H. (1999) Health-related quality of life in pediatric bone marrow transplant survivors: According to whom? *International Journal of Cancer: Supplement*, 12, 46-51.

[81] Patel, H. I. (1985) Analysis of incomplete data in a two-period crossover design with reference to clinical trials. *Biometrika*, 72, 411-418.

[82] Patel, H. I. (1991) Analysis of Incomplete data from a clinical trial with repeated measurements. *Biometrika*, 78, 609-919.

[83] Pepe, M. S. (1992) Inference using surrogate outcome data and a validation sample. *Biometrika*, 79, 2, 355-365.

[84] Pepe, M. S., Reilly, M. and Fleming, T. R. (1994) Auxiliary outcome data and the mean score method. *Journal of Statistical Planning and Inference*, 42, 137-160.

[85] Pocock, S. J. (1979) Allocation of patients to treatments in clinical trials. *Biometrics*, 35, 183-197.

[86] Pocock, S. J., Simon, R. (1975) Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, 31, 103-115.

[87] Reilly, M. and Pepe, M. S. (1995) A mean score method for missing and auxiliary covariate data in regression model. *Biometrika*, 82, 2, 299-314.

[88] Richardson, B. A. and Flack, V. F. (1996) The analysis of incomplete data in the three-period two-treatment cross-over design for clinical trials. *Statistics in Medicine*, 15, 2, 127-143.

[89] Robins, J. M., Rotnizky, A. and Zhao, L. P. (1995) Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90, 106-121.

[90] Rosenberger, W. F. and Lachin, J. M. (1993) The use of response-adaptive designs in clinical trials. *Controlled Clinical Trials*, 14, 471-484.

[91] Rosenberger, W. F. (1995) Adaptive designs in clinical trials: some issues with emphasis on asymptotic inference. *Adaptive Designs.* Institute of Mathematical Statistics Lecture Notes-Monograph Series, Vol. 25.

[92] Rotnizky, A., Robins, J. M. and Scharfstein, D. O. (1998) Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association*, 93, 1321-1339.

[93] Rubin, D. B. (1976) Inference and missing data. *Biometrika*, 63, 581-592.

[94] Rubin, D. B. (1979) Illustrating the use of multiple imputations to handle nonresponse in sample surveys. *Proceedings of 42nd Session of the International Statistical Institute*, Book 2, 517-532.

[95] Rubin, D. B. (1987) *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley.

[96] Rubin, D. B. (1998) Multiple Imputation after 18+ years. *Journal of the American Statistical Association*, 91, 473-489.

[97] Rubin, D. B. and Schenker, N. (1986) Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association*, 81, 366- 374.

[98] Rubin, D. B. and Schenker, N. (1987) Interval estimation from multiple imputed data: A case study using agriculture industry codes. *Journal of Official Statistics*, 3, 375-387.

[99] Rubin, D. B. and Schenker, N. (1991) Multiple imputation in health-care data bases: An overview and some applications. *Statistics in Medicine*, 10, 585-598.

[100] Schafer, J.L. (1997) *Analysis of Incomplete Multivariate Data.* Chapman & Hall.

[101] Schenker, N., Treiman, D. J. and Widman, L. (1993) Analyses of public use decennial census data with multiple imputed industry and occupation codes. *Applied Statistics*, 42, 545-556.

[102] Senn, S. (1993) *Cross-over Trials in Clinical Research.* Chichester: Wiley.

[103] Simon, R. (1977) Adaptive treatment assignment methods and clinical trials. *Biometrics*, 33, 743-749

[104] Simon, R. (1991) A decade of progress in statistical methodology for clinical trials. *Statistics in Medicine*, 10, 1789-1817.

[105] Tanner, M. A. and Wong, W. H. (1987) The calculation of posterior distributions by data augmentation (C/R: p541-550). *Journal of the American Statistical Association*, 82, 528-540.

[106] Taylor, J. M. G., Munoz, A., Bass, S. M., Saah, A., Chmiel, J. S. and Kingsley, L. A. (1990) Estimating the distribution of times from HIV seroconversion to AIDS using multiple imputation. *Statistics in Medicine*, 9, 505-514.

[107] Trenkler, G. and Stahlecker, P. (1996) Dropping variables versus use of proxy variables in linear regression. *Journal of Statistical Planning and Inference*, 50, 65-75.

[108] Wei, L.J. (1979) The generalized Polya's urn design for sequential medical trials. *The Annals of Statistics*, 7, 291-296.

[109] Wei, L. J. and Durham, S. (1978) The randomized play-the-winner rule in medical trials. *Journal of the American Statistical Association*, 73, 840-843.

[110] Wickens, M. R. (1972) A note on the use of proxy variables. *Econometrica*, 40, 759-761.

[111] Williams, E. J. (1949) Experiment designs balanced for the estimation of residual effects. *Australian Journal of Scientific Research*, A2, 351-363

[112] Yao, Q., Wei, L. J. and Hogan J. W. (1998) Analysis of incomplete repeated measurements with dependent censoring times. *Biometrika*, 85, 139-149.

[113] Zelen, M. (1969) Play-the-winner rule and the controlled clinical trial. *Journal of the American Statistical association*, 64, 131-146.